

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 12, 2001, 12:07:44 ; Search time 3358.72 Seconds
(without alignments)
3373.639 Million cell updates/sec

Title: US-09-461-774-1
Perfect score: 1617
Sequence: 1 agcaagcgcgatcgaatcga.....accacggcgcgcgcactga 1617

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 7991742 seqs, 3503743858 residues
Total number of hits satisfying chosen parameters: 15983484

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: em_estba:*
42: em_estfun:*
43: em_esthum1:*

44: em_esthum2:*
45: em_esthum3:*
46: em_esthum4:*
47: em_esthum5:*
48: em_esthum6:*
49: em_esthum7:*
50: em_esthum8:*
51: em_esthum9:*
52: em_esthum10:*
53: em_esthum11:*
54: em_esthum12:*
55: em_esthum13:*
56: em_esthum14:*
57: em_esthum15:*
58: em_esthum16:*
59: em_esthum17:*
60: em_esthum18:*
61: em_esthum19:*
62: em_esthum20:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estov1:*
68: em_estov2:*
69: em_estp11:*
70: em_estp12:*
71: em_estp13:*
72: em_estp14:*
73: em_estp15:*
74: em_estro1:*
75: em_estro2:*
76: em_estro3:*
77: em_estro4:*
78: em_estro5:*
79: em_estro6:*
80: em_estro7:*
81: em_estro8:*
82: em_estro9:*
83: em_estro10:*
84: em_estro11:*
85: em_estro12:*
86: em_estro13:*
87: gb_est41:*
88: gb_est42:*
89: gb_est43:*
90: gb_est44:*
91: gb_est45:*
92: gb_est46:*
93: gb_est47:*
94: gb_est48:*
95: gb_est49:*
96: gb_est50:*
97: gb_est51:*
98: gb_est52:*
99: gb_est53:*
100: gb_est54:*
101: gb_est55:*
102: gb_est56:*
103: gb_est57:*
104: gb_est67:*
105: gb_est68:*
106: gb_est69:*
107: gb_est70:*
108: gb_est71:*
109: gb_est72:*
110: gb_est73:*
111: gb_est74:*
112: em_esthum21:*
113: em_esthum22:*
114: em_esthum23:*
115: em_estom1:*
116: em_estom2:*

117: em_estp16:*
 118: em_estp17:*
 119: em_estp18:*
 120: em_estp19:*
 121: em_estp20:*
 122: em_estp21:*
 123: em_estp22:*
 124: em_estp23:*
 125: em_estp24:*
 126: em_estp25:*
 127: em_estp26:*
 128: em_estp27:*
 129: em_estp28:*
 130: em_estp29:*
 131: em_estp30:*
 132: em_estp31:*
 133: em_estp32:*
 134: em_estp33:*
 135: em_estp34:*
 136: em_estp35:*
 137: em_estp36:*
 138: em_estp37:*
 139: em_estp38:*
 140: em_estp39:*
 141: em_estp40:*
 142: em_estp41:*
 143: em_estp42:*
 144: em_estp43:*
 145: em_estp44:*
 146: em_estp45:*
 147: em_estp46:*
 148: em_estp47:*
 149: em_estp48:*
 150: em_estp49:*
 151: em_estp50:*
 152: em_estp51:*
 153: em_estp52:*
 154: em_estp53:*
 155: em_estp54:*
 156: em_estp55:*
 157: em_estp56:*
 158: em_estp57:*
 159: em_estp58:*
 160: em_estp59:*
 161: em_estp60:*
 162: em_estp61:*
 163: em_estp62:*
 164: em_estp63:*
 165: em_estp64:*
 166: em_estp65:*
 167: em_estp66:*
 168: em_estp67:*
 169: em_estp68:*
 170: em_estp69:*
 171: em_estp70:*
 172: em_estp71:*
 173: em_estp72:*
 174: em_estp73:*
 175: em_estp74:*
 176: em_estp75:*
 177: em_estp76:*
 178: em_estp77:*
 179: em_estp78:*
 180: em_estp79:*
 181: em_estp80:*
 182: em_estp81:*
 183: em_estp82:*
 184: em_estp83:*
 185: em_estp84:*
 186: em_estp85:*
 187: em_estp86:*
 188: em_estp87:*
 189: em_estp88:*

190: gb_gss25:*
 191: gb_gss26:*
 192: gb_gss27:*
 193: gb_gss28:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	195.2	12.1	1101	190	CNS00YXJ	AL096961 Drosophila
C 2	194	12.0	671	135	BE775585	BE775585 MY-04-C-0
C 3	191.8	11.9	579	107	BE344498	BE344498 EST313562
C 4	191.4	11.8	597	107	BE344505	BE344505 EST313569
5	186.8	11.6	637	135	BE777047	BE777047 MY-23-H-0
6	185.6	11.5	635	135	BE776273	BE776273 MY-13-B-0
7	178.4	11.0	740	12	AA820868	AA820868 LD24787.5
8	176.4	10.9	728	14	AA979150	AA979150 LD33429.5
9	175.2	10.8	637	135	BE775334	BE775334 MY-01-F-0
10	173.6	10.7	700	91	AW506070	AW506070 GE1379 G1
11	171.8	10.6	674	16	A1106885	A1106885 GH06135.5
12	170.4	10.5	716	140	D46006	D46006 R1C510372A
13	168.6	10.4	1351	108	BE422296	BE422296 HMM022CF
14	167.8	10.4	645	87	AW224051	AW224051 EST300862
15	167.6	10.4	582	18	A1292742	A1292742 GH15714.5
16	166.4	10.3	616	110	BE583286	BE583286 6-2D-MY P
17	166.4	10.3	624	21	A1517089	A1517089 GH27722.5
18	165.6	10.2	637	87	AW223768	AW223768 EST300579
19	164.4	10.2	819	110	BE643470	BE643470 C-12.8.O1
20	159.4	9.9	513	37	AV620989	AV620989 AV620989
C 21	159.4	9.9	831	111	BE705255	BE705255 SC02.10C0
22	155.2	9.6	632	21	A1517559	A1517559 GH28564.5
23	154.8	9.6	695	106	BE318972	BE318972 NF043C12L
24	154.2	9.5	626	24	A1729076	A1729076 BNGH1125
25	152.8	9.4	736	134	BE053039	BE053039 GA_Ea001
26	150.8	9.3	825	97	AW927061	AW927061 HVSME000
27	149.2	9.2	500	108	BE420380	BE420380 WMS06_A2R
28	148.4	9.2	562	16	A113596	A113596 GH0960.5
29	146.4	9.1	510	36	AV393176	AV393176 AV393176
30	146.4	9.1	668	90	AW398404	AW398404 EST298251
31	143.8	8.9	522	28	AV642726	AV642726 AV642726
32	143.6	8.9	598	26	A1895994	A1895994 EST265437
33	142.6	8.8	660	21	A1546499	A1546499 LD48090.5
34	141.6	8.8	694	96	AW907067	AW907067 EST345099
35	141.2	8.7	709	21	A1486676	A1486676 EST244998
36	141	8.7	434	37	AV640061	AV640061 AV640061
37	141	8.7	445	37	AV636074	AV636074 AV636074
38	140.4	8.7	600	108	BE432172	BE432172 EST398701
39	140.2	8.7	707	6	AA391130	AA391130 LD10228.5
40	139	8.6	902	105	BE231179	BE231179 HVSME001
41	138.2	8.5	519	21	A1533866	A1533866 SDO5838.5
42	138.2	8.5	561	21	A1542174	A1542174 SDO8309.5
43	138.2	8.5	611	21	A1531028	A1531028 SDO1884.5
44	138	8.5	488	36	AV392414	AV392414 AV392414
45	137.8	8.5	640	18	A1296971	A1296971 LP11105.5

ALIGNMENTS

RESULT 1
 CNS00YXJ/c
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS

CNS00YXJ 1101 bp DNA
 Drosophila melanogaster genome survey sequence T7 end of BAC
 BACN01124 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
 AL096961
 AT096961.1 GI:5608572
 GSS.

[illegible]

Db	404	GCACCTCAGACTTGTGTTGAACCCGCCCTAAAGCGTGCCCTCAAGTGTGGCGGTGAAG	345
Qy	823	ggccgctacttcgtgacccgcctaaggcgttctcttgaaagaccitggcggtgtgaacgagt	882
Db	344	GCACCOCGKTTTGGTAGAATATGCAAGGAGGAAATTGTGATGACATGGCCGTGGCCACCGGT	285
Qy	883	ggccacgttgttaaccgccgaagccggcatlgtctgtgcgaag---tgggcttggagtq	939
Db	284	GGAATCKTKTTTGGAGACCAAGCAATATGTGTACCGCTCGAGGATTCAAAGATGAGCGAY	225
Qy	940	cctggctcgcccgcgcgcgtgtgttgcgaagaagacgaacgaagttatgtlgaagcggcgc	999
Db	224	TTTTGACGKTATGGCGAGAGTAAGTACTCTCGAAGAAGCACACATCTCTCTCAAGGCGAAG	165
Qy	1000	ggcacccgcgaagcagcgtgtgcaccaaccgcggcgcaaccaacttgcgttgcagatgcacaagac	1059
Db	164	GGACAGAAAGGCGAGAGGTGGAGAAAAGGCTCGAGGGTCTGCGCGAGCCCATMAAAAACTC	105
Qy	1060	ga-ttcggaattgggaltcgggaaaagcatttgcgcagcgcgtgcgccaacttgacc	1110
Db	104	GACTTCCTCCTACGAGAAAGGAGAAATGACAAGAGCGCTTGCGCCSCCTCTCY	53
RESULT	2		
LOCUS	BE775585	671 bp	mRNA
DEFINITION	MY-04-C-01 Pfinfestansmy Phytophthora infestans cDNA, mRNA sequence.	EST	20-SEP-2000
ACCESSION	BE775585		
VERSION	BE775585.1	GI:10229240	
KEYWORDS	EST.		
SOURCE	potato late blight agent.		
ORGANISM	Phytophthora infestans		
REFERENCE	Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae;		
AUTHORS	Phytophthora.		
TITLE	1 (bases 1 to 671)		
JOURNAL	Kamoun,S., Hraber,P.T., Sobral,B.W.S., Nuss,D. and Govers,F.		
MEDLINE	Initial assessment of gene diversity for the oomycete pathogen		
COMMENT	Phytophthora infestans based on expressed sequences		
	Fungal Genet. Biol. 28 (2), 94-106 (1999)		
	20056376		
FEATURES	Contact: Govers F		
SOURCE	Laboratory of Phytopathology		
	Wageningen University		
	Binnenhaven 9, P.O.Box 8025, 6700 EE, Wageningen, The Netherlands		
	Tel: 31 317 483 138		
	Fax: 31 317 483 412		
	Email: Francine.Govers@medew.fyto.wau.nl.		
	Location/Qualifiers		
	1..671		
	/organism="Phytophthora infestans"		
	/strain="DDR7602, Al mating type"		
	/db.xref="taxon:4787"		
	/clone_lib="PfinfestansMW"		
	/dev_stage="4-week old vegetative, non-sporulating		
	mycelium in synthetic medium"		
	/lab_host="E. coli, strain DH5-alpha"		
	/note="Vector: pSPORF1. Site1: SalI. Site2: NotI. Total		
	RNA was isolated from mycelium of P. infestans DDR7602		
	cultured for 4 weeks in synthetic medium. EST clones were		
	named by their position in the microtiter plate, preceded		
	by the prefix MY (for mycelial) and the successive number		
	of the microtiter plate (e.g. MY-06-A-04)."		
BASE COUNT	136 a	205 c	229 g
ORIGIN	101 t		
Query Match	12.0%; Score 194; DB 135; Length 671;		
Best Local Similarity	60.7%; Pred. No. 1,4e-32;		
Matches	353; Conservative 0; Mismatches 225; Indels 4; Gaps 2;		
Qy	2	gcaagcgtatcgaataacgaacacgcgcgtgcgcacatggaagtgatgcgatygacaagc	61
Db	89	GCAAGGACATCGCGCTTCGCGCGGAGAGTCCGGCCGCTTATCGCTTAAGGCGCCGATCAGC	148

FEATURES	Source
LOCATION/Qualifiers	1..579
ORGANISM="Lycopersicon esculentum"	
CULTIVAR="T4496"	
/db_xref="taxon:4081"	
/clone="cLEY7A23"	
/clone_lib="tomato root, plants pre-anthesis, Cornell University"	

```

/issue_type="root"
/dev_store="plants in pre-anthesis stage"
/notes="vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; supplier: Tanksley; Tissue supplied by Dave Garvin
(USDA-AAS, Ithaca, NY 14850)."
BASE COUNT      109 a      155 c      170 g      145 t
ORIGIN
Query Match      11.9%; Score 191.8; DB 107; Length 579;
Best Local Similarity 65.1%; Pred. No. 4.1e-32;
Matches 283; Conservative 0; Mismatches 153; Indels 0; Gaps 0;

QY 14 aatagagacaaacgcgcgtcgccatggaagtcgtgcataagctggccagaccg 73
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 445 AATTGGTATTTCCTCGCTCTAAGAAAATGCTGTGGTGTCAACGTTCTGGCTGACGCG 386

QY 74 tgcgggtgacgtcttgggcccgcgcgcgcgcgcgcgtgtgtcttgcgaagcgttctggcgac 133
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 385 TAAAGCGACGCTTGCGGCCGGAAGGCCGCTTACAGTGGTACTGGCCAAAGAGCTTCGGCGCCG 326

QY 134 ccacggttaccacagacgagcgctgcacggtgtgcacgttgagatcgcagcttggaagatcccgltg 193
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 325 GCACCATCATCCAAAGGACGGCGCTTCCGTCGCCAAGAAATCGACGTGAAGAAGCGGTTTCG 266

QY 194 aagacttggcgccacagctgtgtgaagtcggtgtgccaaccaagaacgaatgttccggtg 253
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 265 AAAAATGCGGCCCGCCAGCGGTCTCAAGGAAGTTGCTTCCAAAGCCAAACGACGCTGCCGCGTG 206

QY 254 acgagacacacacacacacatcttggcgcaagacactgaataaggcgagcctgaagactag 313
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 205 ACGGCAACCAACCAACCCCTACCGTTCTGGGCTCAGGCACTCTCAACGAAGGCGCTGAAGACCG 146

QY 314 tggccgcgcgcgtcaacccgcgcgcgcgcgcgcgtgggaatcgcgaagcgccgcgacgcg 373
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 145 TGGCTGCCGCGCATGAACCGGATGACCTGAACGAGGCGGTATGCACAAGCGCATCTGCTGCCG 86

QY 374 taatcgagagcgtctgtgcatacgcgcgcgcgcgcgtgtccgcgaagacgcgcgcgcgag 433
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 85 TCGTTGCCGACACTGAAGAACCTGTCCAAAGCCATGCGCGGACTCAAGGCGCATCGGCCAGG 26

QY 434 tggcgcgcgcgtctc 448
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 25 TAGGTACCATATATCT 11

RESULT 4
BE344505/c 597 bp mRNA EST 17-JUL-2000
DEFINITION EST313369 tomato root, plants pre-anthesis, Cornell University
Lycopersicon esculentum cDNA clone cLEYC23, mRNA sequence.
BE344505
ACCESSION BE344505.1 GI:9254006
KEYWORDS EST.
SOURCE Tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids
I; Solanales; Solanaceae; Solanum; Lycopersicon.
1 (bases 1 to 597)
van der Hoeven,R.S., Garvin,D., Matern,A.L., Holt,I.E., Liang,F.,
Upston,J., Hansen,T., Craven,M.B., Bowman,C.L., Ahn,S., Rongling,C.M.,
Fraser,C.M., Martin,G.B., Giovannoni,J.J. and Tanksley,S.D.
Generation of ESTs from tomato root tissue
Unpublished (1999)
CONTACT: David Pritsch
Clemson University Genomics Institute
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 4366
Fax: 864 656 4293
Email: dfrisch@clemson.edu
5 prime sequence.

```

[illegible][illegible]

Db	227	ACTGACCATGGGCCCCAAGGCTCGCATATGTATCATTTGACGATGCTGGGGCTCGCCAA	286
OY	138	ggttaccacagacgagcgttcacggttgacacgltgacatcgagctcgtgaagatccgtttgaaga	197
Db	287	AATCAACCAAGAGGAGGGCGTCTACTGTGGCCAAAGTCATTGAGCTGGAAGGACAAATTTCAGAA	346
OY	198	ctttggcgcccaactggtgtgaagtcggttgacccaagaacaaagatltgtccggtgaagc	257
Db	347	CATCGGTGCAAGCTGGTGCTCAGAGATGTGGCCAAACAAATACCAACGAGAGGCGGTGATGG	406
OY	258	caccacacccgacacacattctgtgacgaagcactgtatcaagagcgccctgaagctagtgtgc	317
Db	407	CACCAACCAAGGCGACCGTTCGTGGCGCCATTGCCAAGGAGGGCTTCGACAAAGATCTC	466
OY	318	cgcgcgagcttcaaccccggtctcgccgtcgagcgtgtgaatctgacgaagcgccgcgcgaagcgt	377
Db	467	CAAGGGCGCCCATATCCCGTGCAGATTTCGTGCGCGTGTCTATGCTGGCGCGTTGAACCGTCAA	526
OY	378	cgaagcgctgctgtgacatcgagccagccgagtgctcgagcaagaacgcgacatcgccaagtgagc	437
Db	527	GGACACACCTTTAAGACCAATCTCCGCCCGCGTTAGCAGCCCGCAGAGATCGCCAGGTGGC	586
OY	438	gacggtgtcccttcgcgcgacgagcagcagc---atcggtacctggtgttgacgaagatgaacaa	494
Db	587	CACCATCTGGCGCAACAGGCTGATCAAGCCATCGGCAATCTCATCAACCGAGGCGCATATGA	646
OY	495	ggtcgccacacagacagcgtgtgtacgatacgaagaatccctccacgcgcgtggcaccgaattga	554
Db	647	GSTGGAGCGCATGCGCGTTATTCACCGTCAAGATGTGGCAAGACCCCTCACCAGATGACTTGA	706
OY	555	gttccacgaaggttggcttc	576
Db	707	GGTATTCGAGCGCATGAAGTTTC	728
RESULT	9		
LOCUS	BE775334	637 bp	mRNA
DEFINITION	MY-01-F-07 PinfestansmY	Phytophthora infestans cDNA, mRNA sequence.	
ACCESSION	BE775334		
VERSION	BE775334.1	GI:10229044	
KEYWORDS	EST.		
SOURCE		potato late blight agent.	
ORGANISM		Phytophthora infestans	
		Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae;	
		Phytophthora.	
REFERENCE	1 (bases 1 to 637)		
AUTHORS	Kamoun, S., Haberer, P.T., Sobral, B.W.S., Nuss, D. and Govers, F.		
TITLE	Initial assessment of gene diversity for the oomycete pathogen		
	Phytophthora infestans based on expressed sequences		
JOURNAL	Fungal Genet. Biol. 28 (2), 94-106 (1999)		
MEDLINE	20056376		
COMMENT		Contact: Govers F	
		Laboratory of Phytopathology	
		Wageningen University	
		Bilthoven 9, P.O. Box 8025, 6700 EE, Wageningen, The Netherlands	
		Tel: 31 317 483 138	
		Fax: 31 317 483 412	
		Email: Francine.Govers@medew.fyto.wau.nl.	
FEATURES		Location/Qualifiers	
source	1..637		
	/organism="Phytophthora infestans"		
	/strain="DDR7602, Al mating type"		
	/db_xref="taxon:4787"		
	/clone_lib="pinfestansmY"		
	/dex_stage="4-week old vegetative, non-sporulating"		
	/mycelium="in synthetic medium"		
	/lab_host="E. coli, strain DH5-alpha"		
	/note="Vector: pSPOR1; Site 1: SalI; Site 2: NotI; Total		
	RNA was isolated from mycelium of P. infestans DDR7602		
	cultured for 4 weeks in synthetic medium. EST clones were		
	named by their position in the microtiter plate, preceded		
	by the prefix MY (for mycelial) and the successive number		

[illegible][illegible]

TITLE Generation of ESTs from tomato fruit tissue
JOURNAL Unpublished (1999)
COMMENT Contact: David Fritsch
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 4366
Fax: 864 656 4293
Email: dfritsch@CLEMSON.EDU

FEATURES
source Location/Qualifiers
1.645
/organism="Lycopersicon esculentum"
/cultivar="7496"
/db_xref="taxon:4081"
/clone_lib="tomato fruit red ripe, TAMU"
/tissue_type="pericarp"
/dev_stage="red ripe (7-20 days post-breaker)"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2: XhoI; Supplier: Glvoanoni; Fruit were tagged at the breaker stage (first sign of lycopene accumulation on the blossom end of the fruit) and harvested 7 days post-breaker (fully red-ripe), 10 days post breaker, and 20 days post-breaker (over-ripe). 20 day fruit which showed external or internal signs of pathogenesis were discarded. Fruit were cut in half and the seeds and locules were discarded prior to freezing the pericarp."

BASE COUNT 205 a 102 c 166 g 172 t
ORIGIN

Query Match 10.4%; Score 167.8; DB 87; Length 645;
Best Local Similarity 53.8%; Pred. No. 7.2e-27;
Matches 346; Conservative 0; Mismatches 297; Indels 0; Gaps 0;

QY 590 tctctgactctctgtaccgactcgataccagcggtgctcgagagcgcttga 649
DB 2 tctctcttcttcttctgttaccgacgacgagaaatgctccgttgatnagacactgttact 61
QY 650 tctctctacacacagacagacatcagctcgtctccgactctgttgcacttctggaagaa 709
DB 62 tgcctactggttgatgataaagatnagacagagatctgttgaatgctcctgagacatg 121
QY 710 ttgcaggaacgggttaagcactactgctgctgagagcggtgagggcgagcgcttgg 769
DB 122 ctatcacaataatggttaccctaatgttattgtctggaagatattgacgacagaaacttgg 181
Y 770 cgaagcgtgctgtaacagcgatlcgcaagcgcttgaaagcggtcgctcaagggcgct 829
DB 182 caactctgtgttgcaataatgacgttagaggtccttgaaaggtcgtgcaactttaaactcctg 241
QY 830 acttcggtgacccgacgtaagcgcttcccttgagagcctgctgaggtgaggtgacgacg 889
DB 242 gttttgttgagcgaataaagcagatattcttatgacatagcaacccttactgagacgactg 301
QY 890 tggtaacaccccgacgacgcatgtgctgaggtgaggtgaggtgaggtgaggtgaggtg 949
DB 302 ttatttagggagacacttggccttacctttaggacaaagcctgacaaaggaattcttaggtcatg 361
QY 950 cccgagcggtggtgctgacgaagagacacagcgttcattgctgagagcggtgagcgacg 1009
DB 362 ctgcttaagatagtgctactaagatgacacacacacacacacacacacacacacacacac 421
QY 1010 aagcggtgacacacgagcgaacacactgctgctgagacacacacacacacacacacacac 1069
DB 422 aagcagcagcaac 481
QY 1070 gggatcggaagac 1129
DB 482 atgataaaggaagac 541
QY 1130 aggttggtgctgacgac 1189

DB 542 aggttgagactcaaacactgaaactgattgaaaggaagaaacttagactgaaacactctc 601
QY 1190 tgcggccgcgaagcgccgctgctgagagagggcactgctcctgg 1222
DB 602 tcactgcaacaaagcagctgttgaggaagactattgtgtttgg 644

RESULT 15
AI292742
LOCUS
DEFINITION GH15714.5prtime GH Drosophila melanogaster head pot2 Drosophila melanogaster cDNA clone GH15714 5prime similar to X99341: Hsp60 PBgn0015245 PID:e254159, mRNA sequence.
AI292742
ACCESSION AI292742.1 GI:3942149
VERSION AI292742
KEYWORDS EST.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 582)
Harvey, D., Hong, L., Evans-Holm, M., Pendleton, J., Su, C., Brokstein, P., Lewis, S. and Rubin, G.M.
BDGP/HMMI Drosophila EST Project
Contact: Harvey, D.
Unpublished (1997)
G. M. Rubin-Molecular and Cell Biology
University of California Berkeley
539 LSA, Berkeley, CA 94720-3200, USA
Fax: 510 643 9947
Email: http://www.fruitfly.org/EST_estefruitfly.berkeley.edu
hit genomic sequence AC004116; hit genomic sequence AC004721
Plate: 157 row: B column: 2
High quality sequence stop: 485.
Location/Qualifiers
1.582
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="GH Drosophila melanogaster head pot2"
/dev_stage="adult"
/lab_host="DHS - alpha"
/note="Organ: head; Vector: pOT2; Site_1: EcoRI; Site_2: XhoI; Sized fractionated cDNAs were directly ligated into pOT2. Plasmid cDNA library."

BASE COUNT 132 a 171 c 173 g 106 t
ORIGIN

Query Match 10.4%; Score 167.6; DB 18; Length 582;
Best Local Similarity 56.7%; Pred. No. 7.9e-27;
Matches 330; Conservative 0; Mismatches 249; Indels 3; Gaps 1;

QY 182 aagatccgttgaagacttggtgagcagcgtgtgtaagtcgttgacacacacacacacacac 241
DB 1 AGGACAGTTCAGACAGCATGCTGCCAAGCTGCTCAGATGTGGCCAAATATCAACG 60
QY 242 atgttgcgttgagac 301
DB 61 AGGAGGCCGGGAGTGCACACACACGCGCACCGCTTGGCTCGGCCCATTTGCCAAGGAGG 120
QY 302 gcttgaagcctagtggtgacccgac 361
DB 121 gcttgcagaaagatctccaaagggcccaatccgctgagatcttgcgtcggtctatgcttg 180
QY 362 ccgacgacggtatcttgagagcgctgctgcatcgcacacgacggtgtccggaagacg 421
DB 181 ccgttgagaccgttcacagac 240
QY 422 gcatcgacaggtggtgacgctgtcctcgacgacgacgacgacgacgacgacgacgacgacg 478

Tue Jan 16 10:17:02 2001

us-09-461-774-1.rst

Page 12

Db	241	AGATGGCCAGTGGCCACCATCTCGGCGCAAGSGTGATTCAGGGCATGGGCAATCTCATCA	300
OY	479	gcgaagcgatgaacaaggtcgcgcacgacagcgtgtgcagcgttcaagaaatcctccacgc	538
Db	301	GCGAGGCGCATMAAGAAGGTGGAGCGCATGGCGTTATCACCGTCAAGATGAGCAAGACC	360
OY	539	tgggacacggatttgagtctccgaaagttatttgcttccaaagagctcttctgcgat	598
Db	361	TCACCGATGATTTGGAGGTTATTCAGGGGCATGAAGTTTCATCCGGGCTTATCTCGCGGT	420
OY	599	acttcgtaccgacttcgataaccagcagcggtgtcctcgagacgcgtltatcctgtac	658
Db	421	ACTTCATCACTCCTCGAAGGGGCCCAAGGGAGGTTCCAAAGATGCCCTCTCTCTGT	480
OY	659	accaagacagaatcagctgcgtctccgatactgtgtgcattcgtgnaaaagttgcagaa	718
Db	481	CCGGAAGAAATATCTCATTCGGTGGCAGAGCATCATTTCCCGCTTGAGACTGGCAATGCGC	540
OY	719	cgggtlaagccactactgatactgtgctctaaagcgtgagggcg	760
Db	541	AGCGCAAGCCGCTGTATCTTTGCCAGGATATTCAGTTGG	582

Search completed: January 12, 2001, 18:01:13
Job time: 21209 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 12, 2001, 12:52:12 ; Search time 153.29 Seconds
(without alignments)
1700.021 Million cell updates/sec

Title: US-09-461-774-1
Perfect score: 1617
Sequence: 1 agcaagcgcgatcgaatacga.....accacggcagcgcgactcga 1617

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 280836 seqs, 80580151 residues
Total number of hits satisfying chosen parameters: 561672

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA :
1 : /cgn2_6/prodata/1/ina/5A.COMB.seq : *
2 : /cgn2_6/prodata/1/ina/5B.COMB.seq : *
3 : /cgn2_6/prodata/1/ina/6.COMB.seq : *
4 : /cgn2_6/prodata/1/ina/PCRTUS.COMB.seq : *
5 : /cgn2_6/prodata/1/ina/backfiles1.seq : *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	741.6	45.9	1620	US-08-461-775-10	Sequence 10, Appl
2	741.6	45.9	1620	US-09-031-606-10	Sequence 10, Appl
3	727.4	45.0	2668	US-08-461-775-11	Sequence 11, Appl
4	727.4	45.0	2668	US-09-031-606-11	Sequence 11, Appl
5	685.6	42.4	1626	US-08-997-080-159	Sequence 159, App
6	685.6	42.4	1626	US-08-997-362-159	Sequence 159, App
7	685.6	42.4	1626	US-09-095-855-159	Sequence 159, App
8	681.8	42.2	1569	US-08-997-080-113	Sequence 113, App
9	681.8	42.2	1569	US-08-997-362-113	Sequence 113, App
10	681.8	42.2	1569	US-09-095-855-113	Sequence 113, App
11	623.8	38.6	1320	US-08-461-775-8	Sequence 8, Appl
12	623.8	38.6	1320	US-09-031-606-8	Sequence 8, Appl
13	609.4	37.7	2167	US-08-461-775-9	Sequence 9, Appl
14	609.4	37.7	2167	US-09-031-606-9	Sequence 9, Appl
15	395.2	24.4	985	US-08-997-080-161	Sequence 161, App
16	395.2	24.4	985	US-08-997-362-161	Sequence 161, App
17	395.2	24.4	985	US-09-095-855-161	Sequence 161, App
18	390.4	24.1	927	US-08-997-080-116	Sequence 116, App
19	390.4	24.1	927	US-08-997-362-116	Sequence 116, App
20	390.4	24.1	927	US-09-095-855-116	Sequence 116, App
21	352.6	21.8	1838	US-08-470-260-7	Sequence 7, Appl
22	352.6	21.8	1838	US-08-471-491-7	Sequence 7, Appl
23	352.6	21.8	1838	US-08-466-662-7	Sequence 7, Appl
24	335	20.7	2284	US-08-467-822-28	Sequence 28, Appl
25	293.4	18.1	647	US-08-997-080-115	Sequence 115, App
26	293.4	18.1	647	US-08-997-362-115	Sequence 115, App
27	293.4	18.1	647	US-09-095-855-115	Sequence 115, App
28	161.4	10.0	342	US-08-473-020A-30	Sequence 30, Appl

29	160.6	9.9	343	2	US-08-473-020A-3	Sequence 3, Appl
30	158.6	9.8	360	2	US-08-997-080-79	Sequence 79, Appl
31	158.6	9.8	360	2	US-08-997-362-79	Sequence 79, Appl
32	158.6	9.8	360	3	US-08-873-970-79	Sequence 79, Appl
33	158.6	9.8	360	3	US-09-095-855-79	Sequence 79, Appl
34	158.2	9.8	343	2	US-08-473-020A-7	Sequence 7, Appl
35	156.6	9.7	342	2	US-08-473-020A-28	Sequence 28, Appl
36	156	9.6	350	1	US-08-105-168B-22	Sequence 22, Appl
37	156	9.6	350	2	US-08-698-948-22	Sequence 22, Appl
38	151.8	9.4	343	2	US-08-473-020A-6	Sequence 6, Appl
39	151	9.3	343	2	US-08-473-020A-4	Sequence 4, Appl
40	150.2	9.3	342	2	US-08-473-020A-31	Sequence 31, Appl
41	147	9.1	343	2	US-08-473-020A-2	Sequence 2, Appl
42	146.2	9.0	343	2	US-08-473-020A-5	Sequence 5, Appl
43	144.6	8.9	343	2	US-08-473-020A-8	Sequence 8, Appl
44	143	8.8	342	2	US-08-473-020A-29	Sequence 29, Appl
45	142.4	8.8	337	2	US-08-997-080-77	Sequence 77, Appl

ALIGNMENTS

RESULT 1
US-08-461-775-10
; Sequence 10, Application US/08461775
; Patent No. 5858773
; GENERAL INFORMATION:
; APPLICANT: MAZODIER, Philippe
; APPLICANT: GUGLIEMI, Gerard
; TITLE OF INVENTION: REGULATORY NUCLEOTIDE SEQUENCE OF THE
; TITLE OF INVENTION: INITIATION OF TRANSCRIPTION
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: George Mason Bldg., Washington & Prince Sts.
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/461,775
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/050,313
; FILING DATE: 10-MAY-1993
; APPLICATION NUMBER: FR 9011186
; FILING DATE: 10-SEP-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Crane-Feunty, Sharon E
; REGISTRATION NUMBER: 36,113
; REFERENCE/DOCKET NUMBER: 010830-035
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1620 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1620
; US-08-461-775-10

COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/461,775
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/050,313
FILING DATE: 10-MAY-1993
APPLICATION NUMBER: FR 9011186
FILING DATE: 10-SEP-1990
ATTORNEY/AGENT INFORMATION:
NAME: Crane-Feury, Sharon E
REGISTRATION NUMBER: 36,113
REFERENCE/DOCKET NUMBER: 010830-035
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 2668 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-461-775-11

Query Match 45.0%; Score 727.4; DB 2; Length 2668;
Best Local Similarity 67.3%; Pred. No. 1.1e-144;
Matches 1092; Conservative 0; Mismatches 516; Indels 15; Gaps 4;

QY 4 aagctgacgaatacgaacacgcgcgtcgcgcacatgagctgagtcgagcaagatc 63
DB 855 AAGATTTCGAATTCGACGAGGAGCGCCGTCGCCCTTGAGCGGGCGTGAACGAG 914
QY 64 gccgaacccgtgagctgagctgagctgagctgagctgagctgagctgagctg 123
DB 915 GCCGACACCGTCGACGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 974
QY 124 ttgacgac 183
DB 975 TTGCGCGCCGACCATACACACACACACACACACACACACACACACACACAC 1034
QY 184 gatccgttgaagacttggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 243
DB 1035 GACCGGTGAGAACCTGGCGGCCGACGTCGTCGTCGTCGTCGTCGTCGTCGTC 1094
QY 244 gttgcggttgaacgacacacacacacacacacacacacacacacacacacac 303
DB 1095 ATCGGGGTGAGACGACACACACACACACACACACACACACACACACACAC 1154
QY 304 ctgagagcagtgagc 363
DB 1155 CTGGCCAAAGTCG 1214
QY 364 gccacgcggtatctgagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 423
DB 1215 GTGCGCGCGCGTCGCGCGAGCTGCTCGACACCGCGCGCGCGCGCGCGCGCG 1274
QY 424 atcgcgcagtgagc 483
DB 1275 ATCG 1334
QY 484 gcatgacacagtgatcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 543
DB 1335 GCGATGACAGGTCGCGAAGACGCTGTATCAACGTCGAGAGTCCACACACTTCG 1394

QY 544 accgagttgaggttaccgaagatattggtctccacaaaggcttcttctgacatctc 603
DB 1395 GTGACACTGGAATTCACCGAGGAGCATGGCTTCGACAAAGGGTACTCTCCCGCATG 1454
QY 604 gttaccgacttcgaataaccagcagcgttctcgagagcgcgttgaatctctacacaa 663
DB 1455 GTGACCGACAGAGACCGATGAGAGCGCTCTCCGACGACCGCGTACTCTGATCCAC 1514
QY 664 gacaagatcagctctccgcgtcttgcgttgcgttgcgttgcgttgcgttgcgtt 723
DB 1515 GGCAGATCGGTTTCATACAGACCTGCTCCGCTGCTGAGAAAGGTATCATCAGCGG 1574
QY 724 -----aagcactactgactgcttgcgttgcgttgcgttgcgttgcgttgcgtt 777
DB 1575 GCGTCCAAAGCGCGCTGATATCGCCGAGAGACGTCGAGGGGAGAGCGCTGACCC 1634
QY 778 gttgtaacgcgacttcgaacagcgttgaagcgcgttcgcgttgcgttgcgttgcgt 837
DB 1635 GTGTCACAAAGATCGCGGACGCTTCAACGCGCGCTGCTCAAGGCGCGCGCTTCG 1694
QY 838 gaccgcgtaagcgttctcttgcgttgcgttgcgttgcgttgcgttgcgttgcgtt 897
DB 1695 GACCGCCGACAGCATGCTGCGGACATGGCCACCTTCACCGGTGACACCTATCCGC 1754
QY 898 cccgacgcgcgcatctgtctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 957
DB 1755 GAGGAGGTGCGGCTCAAGCTGACACGCGCGGTGTCGAGGTGTCGACCGCGCGCG 1814
QY 958 gttgtgttcgaagagacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1017
DB 1815 GTACACGTCACCAAGAGACGACGACGACGACGACGACGACGACGACGACGACG 1874
QY 1018 gccaacgc 1077
DB 1875 CAGGCGCGCGCTGCGCCATACATCAAGCCGACGATGAGTGCACCGACTCGGACCG 1934
QY 1078 gaaagacttgc 1137
DB 1935 GAGAGGTTCGAGAGACGCGCTGCGCAAGCTGGCGCGCGCGCTGCTGATCGCGTCCG 1994
QY 1138 gccgcacacgc 1197
DB 1995 GCGGCGACCGAGGTGACTGACGACGACGACGACGACGACGACGACGACGACGACG 2054
QY 1198 gccaaagc 1257
DB 2055 ACCCGCGCGCGGTGAGAGAGGACATCTCTCGGTGCTCCGCTCGCTGCTCAC--- 2111
QY 1258 gccgc 1317
DB 2112 GCGGTCAAGGTCGTGACGACGACGACGACGACGACGACGACGACGACGACGACG 2168
QY 1318 gaagtgtcttcgaagcccttgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1377
DB 2169 GCGGTGTCGCTC 2228
QY 1378 gaacgctcgttgcgttgaacaggttcagcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1437
DB 2229 GAGGCTTACGTCATACCAACCAAGTGGCGGAGCTCGACAAAGGGCACGAGGCTTCA 2288
QY 1438 aacacccctgaactatgtgacttgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1497
DB 2289 GCGACCGCGGAGTACGAGGACGACTGTCAAAGCGCGCGTCATCGACCGGTCGAAGTCA 2348
QY 1498 aggtcgcgcgttgc 1557
DB 2349 CGCTTCGCGCTGAGAAAGCGCGCGCTCCATCGCTCTGCTGACACGACGACGCTG 2408
QY 1558 gttgtcgaacagc 1614
DB 2409 GTCTGTGAGAAAGCGCGCGAGAGAGGCGCGGCTACGCTACGCGGACGACGACG 2468
QY 1615 tga 1617

Db 2469 TGA 2471

|||||

RESULT 4

US-09-031-606-11

; Sequence 11, Application US/09031606

; Patent No. 6153404

; GENERAL INFORMATION:

; APPLICANT: MAZODIER, Philippe

; APPLICANT: GUGLIEMI, Gerard

; TITLE OF INVENTION: REGULATORY NUCLEOTIDE SEQUENCE OF THE

; TITLE OF INVENTION: INITIATION OF TRANSCRIPTION

; NUMBER OF SEQUENCES: 15

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Burns, Doane, Swecker & Mathis

; STREET: George Mason Bldg., Washington & Prince Sts.

; CITY: Alexandria

; STATE: Virginia

; COUNTRY: United States

; ZIP: 22313-1404

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/031,606

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/050,313

; FILING DATE: 10-MAY-1993

; APPLICATION NUMBER: FR 9011186

; FILING DATE: 10-SEP-1990

; ATTORNEY/AGENT INFORMATION:

; NAME: Crane-Feury, Sharon E

; REGISTRATION NUMBER: 36,113

; REFERENCE/DOCKET NUMBER: 010830-035

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (703) 836-6620

; TELEFAX: (703) 836-2021

; INFORMATION FOR SEQ ID NO: 11:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2668 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

US-09-031-606-11

Query Match 45.0%; Score 727.4; DB 3; Length 2668;

Best Local Similarity 67.3%; Pred. No. 1.1e-144;

Matches 1092; Conservative 0; Mismatches 516; Indels 15; Gaps 4;

QY 4 aagctatcgaatacgaacgaacgcgcgtcgccatgtgagatcgtgcatggaacgtg 63

Db 855 AAGATTGTGAAGTTGCGACGAGGAGCGCCGTCGCGCCCTTGAGCGCGCGCTGAGCACACCTG 914

QY 64 gccacacccgtgagcggtgagcgctgagcgcgccgagcgtggtgtggtgagcgaagcg 123

Db 915 GCCGACACCCCTCAAGGTGACCATTCGCGCCCAAGGCGCGCAACGTTCGTCATCGACAAAG 974

QY 124 ttgtgcgagcccaacggtttaccaaagcgcgtcaggttgacggtgacgtgagatcgtgaa 183

Db 975 TTGCGCGCCCGACCATCAACCAAGAGCGGTCACATCGCCCGTGAGTCGAGTGGGAG 1034

QY 184 gaccggttgaagaacttgagcgcccaagctggtgtaagtccgttgagcacaagaacacgat 243

Db 1035 GACCCTGTACAGAACTCTGGCGCCACCTCGTCAAGAGAGTGGGAGCAAGACCAAGAC 1094

QY 244 gttagcggtgagcgacacacccgaacacatcttgtagcaggaacatcgtatcaaggcgcg 303

Db 1095 ATCGGGGTGACGSGCACACCACCACCGGACCGTGTGGCCCAAGCGCTGTGTCGCGAGGCG 1154

QY 304 ctgaagctagttagccgcgcgcgtlcaaacccggttcgcgcgtcgtggtggaatcgcgaagcc 363

Db 1155 CTGGCAACGTCGCGCGCGGCGCTCCCGCGCCCTGGAAGAGGGGCAATCGAGCGCGCC 1214

QY 364 gccagcggtatctgagcgctgtgcatcggcacaagccggtgtccgcgaagaacccgc 423

Db 1215 GTCCGCCCGCTTCGCGGAGCTCTCGACACCGCGCGCGGATCGAGCAAGATTCGCGAC 1274

QY 424 atcgcgaggttgagcgaggtgtccgcgcgcgcgaagagaatggtgtgacccgtgttgaa 483

Db 1275 ATCGCCCGCGTCCGCGGCTCTCCGCGAGGAGGAGGCTGCGGAGCTATCGCCGAG 1334

QY 484 gcgaatgaacaagtcgcgcgcgcgaagcaggtgtgcaagcgaagatccctccacgtggc 543

Db 1335 GCGATGAGACAAGTGGGCAAGAGGAGGTGATCACTCGAGAGTCCACACCTTCGGT 1394

QY 544 aacgagttggaagttcaccaagaglatgtgcttccaaaggctctgtgcgacatc 603

Db 1395 GTCGACCTGAGACTTCACCGAGGCGCATGCGCTTGACAAAGGCTTACCTGTCGCTACATG 1454

QY 604 gttaacgacttgaatacgaacgaagcggtgtcgaagagcggttgatccctccacaccaa 663

Db 1455 GTGACCGACCAAGAGCGCTTATGAGAGCGGCTCTGACGACCCGTAATCTGTATCCACAG 1514

QY 664 gacaagaatcagctcgtctccgcatctgttccattgtctgtaaaagtttgcgaagacggt 723

Db 1515 GCGAAGATGCGTGTGATCCAGAGACCTGCTCCGCTGTGAGAAAGTCATCCAGGCGGCT 1574

QY 724 -----aagcactactgactgctgtgctgaaagcgttgagagcgaaagcttgagcagctg 777

Db 1575 GGTCTCCAAAGCGCTGTGATCATCGCGAGAGCGTCCAGGCGCGCGCTGTGACCCCTG 1634

QY 778 gtctgaacgcgattgcaagacgttgaaagcggttgcaagggcgctactcgtg 837

Db 1635 GTGTCAACAAGATCCGCGGACGTTTACGCCGTCCGCTCAAGGCGCGCGCTTCGGT 1694

QY 838 gaaccgcgtgaagcgcttctcttgaggaacttgagcggttggtgacggttgccagtgtgtaac 897

Db 1695 GACGCGCGCAAGGAGGATGTCTGGGAGATGAGCCACCTCAACCGGTGCGACGTCATCGCC 1754

QY 898 ccgagcgcgcatctgtgctgc 957

Db 1755 GAGGAGGTGCGCTCAAGCTGACACAGCGCGGCTGTGAGCTGTGGGCGACCGCGCGCGC 1814

QY 958 gtgtgttcaagcaagagacacgctatctgtgagcgcgcgcgcaaccggaagaagcggtg 1017

Db 1815 GTCAACGTTCAACAGAGACGACGACACCATGTGAGCTTGAGAAAGAGACGCGAGAGCTG 1874

QY 1018 gccaaaccggtggaaccactgtgcgtccgaagatcgaaacgaagcgtatcttgatcgcg 1077

Db 1875 CAGGCGCGCGTCCGCCAGATCAAGCGCGGAGATGAGTGCACGATCGAGTGGAGACCG 1934

QY 1078 gaaaagcttgagagcggtgtgccaacgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1137

Db 1935 GAGAAAGCTTCAGAGAGCGCTTCGCCAAGCTTGCGCGCGCGCTTCGCGATTCGCGTGGC 1994

QY 1138 gccgcacacgaacacgcgcacccaaggaagcaaggaagcgttgagatgtggtgcgcgc 1197

Db 1995 GCGGCCACCAAGTTCGAGCTGAAAGAGCGCAAGACCGTGTGAGAGCGCGCATTCGCGC 2054

QY 1198 gccaaagccgcggttcgagagagcagatcgcgtgtgtgaggaagcctcgcacacacaa 1257

Db 2055 ACCGCGCGCGCGGTCGAGAGGAGGAGATGCTCGGAGTGTGCGGCTGCGGTGTCAC --- 2111

QY 1258 gccgcgaagcgcttgacaggaactgctgtgctgtacgcgcgttgacgaagttccctggtgtc 1317

Db 2112 GCGGTCAAGTCTCTGAGCAACACTTCG---GCCGCGACCGCGACGAGGCGCACCGGCTTC 2168

QY 1318 gacgtgtcttcgaagcccttcgcgcgcgtgtgttcgagatcgccgaacgcttgctgtg 1377

Db 2169 GCGGTCGCCGCGCGCGCGCTGCGATGCGCGGAGAACCGCGCTC 2228
Oy 1378 gacgcctgctgtgtcaacaaggctcaagctacccgagcgatgctgaagct 1437
Db 2229 GAGGCGCTACGTATATACCAACCAAGGTGGCGAGCTCGACAAAGGCCAGGCTTCAACGCG 2288
Oy 1438 aacaccctgaagctatgtgtactgtccgctgaagcgatcatcgaacccggtlcaagtgact 1497
Db 2289 GCCACCGCGAGTACGCGGACCTGTGTCAAGCGCGCTCATGACCGGTCGAAGGTCAAC 2348
Oy 1488 agtgcgaggtgtgtgaagcgctcatcggttgcgagatgtgtatccacacgagagcgctc 1557
Db 2349 GCGTCCGCGCTGTGAGAACGCGGCTTCATCGCTCCCTGTGACGACGACCGCTG 2408
Oy 1558 gtgtgcgaagccgagcgaagcgaagaatcagaacatcac---caagcgacgagcac 1614
Db 2409 GTCGTGAGAACCGCGCGAGAGAGAGAGCCGCGGTGACGTCACGCGGACAGCCAC 2468
Oy 1615 tga 1617
2469 TGA 2471

RESULT 5
US-08-997-080-159
; Sequence 159, Application US/08997080
; Patent No. 5968524
; GENERAL INFORMATION:
; APPLICANT: WATSON, JAMES D.
; APPLICANT: TAN, PAUL E.J.
; TITLE OF INVENTION: METHODS AND COMPOUNDS FOR THE TREATMENT OF IMMUNOLOGICALLY-
; NUMBER OF SEQUENCES: 194
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Law Offices of Ann W. Speckman
; STREET: 2601 Elliott Avenue, Suite 4185
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98121
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/997,080
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Sleath, Janet
; REGISTRATION NUMBER: 37,007
; REFERENCE/DOCKET NUMBER: 11000.1007
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-269-0565
; TELEFAX: 206-269-0563
; TELEX:
; INFORMATION FOR SEQ ID NO: 159:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1626 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-997-080-159

Query Match 42.4% Score 685.6; DB 2: Length 1626;
Best Local Similarity 65.6%; Pred. No. 6.3e-136;
Matches 1035; Conservative 0; Mismatches 534; Indels 9; Gaps 2;

Oy 3 caagctgatacatagacgaacacgagcgtcgcgcacatgtagtgcgcatgagacaagct 62

Db 6 CAAGACAATTGCGTATGACGAAGAGGCCCGCGTGGCTCGAGCGGGGCTTCACGCCCT 65
Oy 63 ggcgcgaacccggtgagtgagcgttggtggtcgagcgagcgagctgtgtgtgtgccaagc 122
Db 66 GCGACAGCGCTTAAGGTGACGTGTGGCCGAAGGTGCAACGTCGTGTCGAGACAGAA 125
Oy 123 gtttgagcgaacccaggttaacgaagcggttcaagctgtgcaagtgagatcagctga 182
Db 126 GTGGGGCGCCCGACGATACCAAGATGCTGTGTCATCGTCGAAGAGATGATGACTTGA 185
Oy 183 agatccgtttgaagacttggtggtccagctgtgtgaagtctgtgtgcaacgaagacaaga 242
Db 186 GACCCGTAACGAGATGCGGCTGTGACTGTGTAAGAGGTCCGCAAGAACGACGACGA 245
Oy 243 tgttcggtgaagcagac 302
Db 246 GTCGCGGGGCGACGCGAC 305
Oy 303 cctgaggtgagtgtgagcgcggtgtcaacccggtcgcgtcgcgtgtggaatcgcgaagc 362
Db 306 COTGGCAACGTGCGAGCGCGGCGCCACCGCTGCGGCTCAAGCGTGGCATGTGACAAAGC 365
Oy 363 ggcgcagcgtgtatcgaagcgctgtgtgcatcggcgaagccggtgtcgggaagaccg 422
Db 366 TGTGAGGCTGTCAACCGACCGCTGCTGTAAGTCGCGCAAGGAGTGTGACACCAAGACA 425
Oy 423 catcgcgaagtgagcgaaggtgtcctcgcgcgaagcgaagcgaatcgtgtgacctgtgtgga 482
Db 426 GATTTCTCCACCGCGGGGATTTCCGCGGCGACACCCAGATGGCGGAGCTATGACCGCA 485
Oy 483 agcgaatgaagcgtgcgcgaagcagcgtgtgcaagctcaaggaatcctcgaagctgag 542
Db 486 GGCATGTGACAAAGTCCGCGCAACGAGGTGTCTATCACCTCGAGAGTGTGAAACACTTGG 545
Oy 543 caaccgagtgagtgatcgaagcgaaggtatgtgtcctcacaagaagcttctgtgcgaact 602
Db 546 COTGACGCTGAGCTCACCGACCGAGGATATGCGCTTCACAAAGGCTATCTCGGGTACTT 605
Oy 603 ggttaccgaacttgataaccgaagcgtgtgtcgaaggaagcgtgtgtcgtctacacca 662
Db 606 GGTGACCGACGCGCGACCGCAGGAAGCCGCTGTGAGATCCCTACATCTCTGTGTAG 665
Oy 663 agacaagatcagctgcgtcccgatcgtgttcgaatgtgtgaaagtgtaaggaagcgg 722
Db 666 CTCGAAGGTGTGACCGCTCAAGGATCTCTCCGCTGCTGGAAGAGTATTCAGGCCGG 725
Oy 723 taagcactactgactgtgtgtgaagcgtgtggaagcgtgtgtgcgaagctgtgtcgt 782
Db 726 CAAGCGCTGCTGATCATCTGCGGAGAGCTGAGGGGCGAGGCCCTGTCCACGCTGTGT 785
Oy 783 caagcgaatcgaagcagtgatgaagcgtgtgcaagcgtgtgtgtgtgtgtgtgtgtgt 842
Db 786 CAACCAAGATCCGCGACCTTCAAGTCCGCTCAAGGCTCCGGGCTTGGGTGACCG 845
Oy 843 ccgtaaagcgttcccttgaggaaccttgagcggtgtgtgaagcgtgtgcgaagcccca 902
Db 846 CCGCAAGCGATGCTGCAAGGATGCGCATCTCTACCGGTGTGAGTGTGTGACGGAAG 905
Oy 903 ggcgcgcatgtgtgtcgaagcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 962
Db 906 AGTCGCGCTGTCCCTGAGAGCCGCGAGCTGCTGCTGTGGGCGAGGCCCAAGGTGCT 965
Oy 963 ggtcagcaaggaagcagcagcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1022
Db 966 GGTACCAAGAGAGAGAGACACCATCTGAGAGGCTGCGGCGATTCGATTCATTCACGCGG 1025
Oy 1023 ccgagcgaacacactgtgtgcgaagatcgaagcgaagcgaagcgaatgtgtgtgtgtgtgt 1082
Db 1026 CCGGCTGCTGATGATCCGCGCGAGATGAGAAACGCACTTCGATTCGACCGCGAGAA 1085
Oy 1083 gtttgcgaagcgtgtgcgaagcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1142

[illegible]

```

1      COMPUTER READABLE FORM:
2      MEDIUM TYPE: Diskette
3      COMPUTER: IBM Compatible
4      OPERATING SYSTEM: DOS
5      SOFTWARE: FASTSEQ for Windows Version 2.0
6      CURRENT APPLICATION DATA:
7      APPLICATION NUMBER: US/09/095,855
8      FILING DATE:
9
10     CLASSIFICATION:
11
12     PRIOR APPLICATION DATA:
13     APPLICATION NUMBER: 08/705,347
14     FILING DATE: 29-AUG-1996
15     APPLICATION NUMBER: 08/873,970
16     FILING DATE: 12-JUN-1997
17     APPLICATION NUMBER: 08/997,362
18     FILING DATE: 23-DEC-1997
19
20     ATTORNEY/AGENT INFORMATION:
21
22     NAME: Sleath, Janet
23     REGISTRATION NUMBER: 37,007
24     REFERENCE/DOCKET NUMBER: 11000,1002c3
25     TELECOMMUNICATION INFORMATION:
26     TELEPHONE: 206-269-0565
27     TELEFAX: 206-269-0563
28
29     TELEX:
30
31     INFORMATION FOR SEQ ID NO: 159:
32
33     SEQUENCE CHARACTERISTICS:
34     LENGTH: 1626 base pairs
35     TYPE: nucleic acid
36     STRANDEDNESS: single
37     TOPOLOGY: linear
38
39     US-09-095-855-159

```

	Query Match	42.4%	Score	685.6:	DB 3:	Length	1626:	
	Best Local Similarity	65.6%:	Pred. No.	6.3e-136:				
	Matches 1035:	Conservative	0:	Mismatches	534:	Indels	9:	Gaps
OY	3 caagctgatacgaatcacgaacaccgcgcttcgcgccatcgagggttccgatgcatgacaagt	62						
Dd								
Dd	6 CAAGACAAATGTCGTMTGAACGAGAAGGCCCGGGCTTGBCCTCGAAGGGGGCCTCAAGCCCTT	65						
OY	63 ggcgcacacgcgttgccgggttgaaegctttgggcccgcgcggccgcgaatttgtttgttcgccaaygc	122						
Dd								
Dd	66 CCCAACACGCCTGTAAGAGGTGACTTTGGGGCCCCGAAGGGTCCCACAGTCGTGCTCGAAGAA	125						
OY	123 gtlttcggcaaccaacggtttaaccaagaacggcgttaacggytgcgaacttgaaatcgaactga	182						
Dd								
Dd	126 GTGGGGGGCCCCCACGATCATCCAAAGATGTTGTCATTGCCCAAGAGATCGAGCTGGA	185						
OY	183 agatccglttgaagaacttgggcgccccagcttgg tgaagtcgttcgtgcaccaaagacaaaga	242						
Dd								
Dd	186 GGACCCCGTAGAGAAATGGCGGTGAGCTGGTCAAAGAGGTTCGCCAAGAAGACCGACGA	245						
OY	243 ttgttcggttgtagaacaccacacccgcaacacctttgctgcgaaggaactyaatcaagygcg	302						
Dd								
Dd	246 CGTCCCGGGCGACGGCACACCACCGCCACCGTCGTCTGACGGCTCTGGTCTGGCGAAG	305						
OY	303 ccttgaggtctaagtyggccgcccgcggttaaaccggctgcgcctccgcgcgttgaaatctgsgaacg	362						
Dd								
Dd	306 CCTGGCGCAAGTCGACGCCGGCGGCCAACCCGCTCGGCTCTCAAGACGTGGCATCGAGNAGGC	365						
OY	363 gcgcacgcggttatctcgaagcgttgttgcatcgcgcgaacacccgcggttgttccgysaaagaccg	422						
Dd								
Dd	366 TGTTCGAGGGTGTACACCACATGTCGTCTGTAAGTGGGCCAAGAGAGTTCGAGACCAAGAGCA	425						
OY	423 catcgcgcaagttgycgaacgttgtccttcgcgcgaagacagaalctcgtatgacvtggtttgcyga	482						
Dd								
Dd	426 GATTTCTGCGACCGCGGGGATTTCCGCGGGCACACCCAGATCGGCGAGACTCATCGCCA	485						
OY	483 agcgaatgaacaaggttcggnccacgaacagcgttggttgagggttcgaagaaatcttcacagcttgg	542						
Dd								
Dd	486 GCCCATGTGCAAGTGTGGCGACAAGAGGTTATCAACCGTCGAGAGATCAACACCTTCGG	545						

[illegible]

QY	1443	ccctgacatctatgctgacttggccgcgtctgacggcgtctcatcgcgaccccgctcaaggctgagctacgtc	1502
Db	1437	CGGTGATGATACGAAGGACCTGCTCTAAGGCGGGCGTGCAGCACCGGTGAAGGTATCACCCCGCTC	1496
QY	1503	ggcgggtgttgaaacgcgtcatcgtgtgccggaatgctacaccacaggaagcgtctcgtcgtc	1562
Db	1497	GAGCGTGTGAGAACCGCGGCGTTCATCGCGGCGCTGTGTTCTTCACTACCAGAGGCGTCGTGTCG	1556
QY	1563	cgacaagccgg	1573
Db	1557	CGACACGCGG	1567

RESULT

```

US-08-997-362-113
: Sequence 113, Application US/08997362
: Patent No. 5985287
:
: GENERAL INFORMATION:
:
: APPLICANT: Tan, Paul
:
: APPLICANT: Hiyama, Jun
:
: APPLICANT: Visser, Elizabeth
:
: APPLICANT: Skinner, Margot
:
: APPLICANT: Scott, Linda
:
: APPLICANT: Prestidge, Ross
:
: TITLE OF INVENTION: COMPOUNDS AND METHODS FOR
:
: TITLE OF INVENTION: TREATMENT AND DIAGNOSIS OF MYCOBACTERIAL INFECTIONS
:
: NUMBER OF SEQUENCES: 194
:
: CORRESPONDENCE ADDRESS:
:

```

```

1 ADDRESS:  Law Offices of Ann W. Spekman
2 STREET:  2601 Elliott Avenue, Suite 4185
3 CITY:  Seattle
4 STATE:  WA
5 COUNTRY:  USA
6 ZIP:  98121
7
8 COMPUTER READABLE FORM:
9
10 MEDIUM TYPE:  Diskette
11
12 COMPUTER:  IBM Compatible
13
14 OPERATING SYSTEM:  DOS
15
16 SOFTWARE:  FastSEO for Windows Version 2.0
17
18 CURRENT APPLICATION DATA:
19
20 APPLICATION NUMBER:  US/08/997,362
21

```

? FILING DATE: ?
 ? CLASSIFICATION: ?
 ? PRIOR APPLICATION DATA: ?
 ? APPLICATION NUMBER: U.S. Patent Application No. 5985287 08/873,970 ?
 ? FILING DATE: June 12, 1997 ?
 ? APPLICATION NUMBER: U.S. Patent Application No. 5985287 08/705,347 ?
 ? FILING DATE: August 29, 1996 ?
 ? ATTORNEY/AGENT INFORMATION: ?

NAME: Sleath, Janet
REGISTRATION NUMBER: 37,007
REFERENCE/DOCKET NUMBER: 11000.1002c
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-269-0565
TELEFAX: 206-269-0563
TELEAX:

```

: INFORMATION FOR SEQ ID NO: 113
:
: SEQUENCE CHARACTERISTICS:
:   LENGTH: 1569 base pairs
:   TYPE: nucleic acid
:   STRANDEDNESS: single
:   TOPOLOGY: linear
:
: MOLECULE TYPE: Genomic DNA
US-08-097-362-113

```

Query Match	42.2%	Score 681.8;	DB 2;	Length 1569;	
Best Local Similarity	65.6%;	Pred. No. 3.9e-135;			
Matches 1030;	Conservative 0;	Mismatches 532;	Indels 9;	Gaps 2;	

Oy 3 caagctgatcgatacgcgacgaaccgcgcgttcgcgccatlgagaagtccgcatlbgacaagct 62
||| | | | ||| | | | | | | | | | | | |
Db 6 CAAGCAATTTGCGTGTACGAAGAAGGCCCGCCGGCTCGAGCGAGGAGCCTTCACGCCCT 65

QY	63	ggagcaacccgtgacggtgaacgcgtggggcccgccgagccgcgcatgttgcttgccaagcg	122
Db	66	cgcagaccctttaaaggctgacgtttggccccgaagcgtccgaacgctgctgtcagacaagaa	125
QY	123	gtttgcygagaccaccggttaccaaacgagcgttcaacgtgtgacatgtgagatctgacgttga	182
Db	126	gtggggcccccctccagatcacaacagatgggtgttcattgcgccaagaagatcagcttgga	185
QY	183	agatccgtttgaagaacttgggcgcgcacgctgtgtgaaagtcgtgtgccaaccaagaacca	242
Db	186	ggaccctgtacagagaagatcgccgtgagctggtctaaagagagtgcccaagaagaccacaga	245
QY	243	tgttcggttgaagcacaaccaaccgacaacatcttggcgcagcactgtgacaaagcgcg	302
Db	246	cgtgcggggccacggcagacacacacccgaccctgcttgccttcagagcttctgttgcgcaag	305
QY	303	ccctgaagctagttgacgcgcgagtgcaaacccggtctgcgtctcgcytggaaatcgcgaagc	362
Db	306	cctgcgcgaacctccacgcggccgccaacccgcttcgcctcaaacctgtggcttcagaaagcg	365
QY	363	cgcccaacggttatctgaaggcgctgcgtgcgtatctggccacgcggtgttccgcaaacgcg	422
Db	366	ttctgagagcttccaccagtcgctgctgaagtggcccaagagagctgagacacaaagagca	425
QY	423	catcgccaaagttgcgcagcgtgttctctcgccgacagacaaatcgtgtgacctgtgttgca	482
Db	426	gattttgtccaccccgccggagattttcccgccggcgacacccaatcggcgagctattcccgca	485
QY	483	agcgatgaacaaggttcgycgaacagacgctgtgtcaagcgttaagaaatcctccaagcttgg	542
Db	486	ggcccatggacaaaggtccggcaacgaggggtgtcatctacgcctggaagagttgcacaccttcg	545
QY	543	caccagatlgagatgtcaacgaaggtatgtgtctccaaaggtcttlttgcatcact	602
Db	546	cctgcagctgcagcttccacccaggggtatggccttgcacaaaggcttaacatctcgggattactt	605
QY	603	cgttacgcgaattctgataaaccaagcagcggtgtctcgaaagcgcgttgcactctgacacaa	662
Db	606	cgtgacgcgaccccgagcgccagaggaagcctgtctgaagagatccttcaatctcgtgtgacg	665
QY	663	agacaagaatacaagctcggttccgaatcgtgtgcatgtgtgaaanaagttgcaagaaacgg	722
Db	666	ctccaaagctgtgcacctctcaaaaggtgtgtgtccgctgtctggaacaaagctatccaaagcccg	725
QY	723	taagccactactgtatcgtgcgtgaagacgtgtgaagggcgaaagcgtttgacgacgtgtgt	782
Db	726	caaacccgctgtatcatccgcggagacgctcgaagggagagccctgtccacgcgtggatgt	785
QY	783	caacgcgaatttcgcaagaacgcttgaagcgcgttcgcgttcaagggcgtaacttcgtgtacg	842
Db	786	caacaacaatccgcgcgcacctttcaaaagttccgttcgctgaacgctccgggtgtcggtgtacgc	845
QY	843	ccgtaagacgtcttccttaagaaacctgtgacgtgtgtgtgaacggtgtgtgcacgttgtaacc	902
Db	846	ccgcaagggcattgctgtgaagacatggccatcttccaccggtgtcagtgctgacgcaag	905
QY	903	cgccgcgaattgtgctgcgcgaagtgtagcgttgaaggtgtgtgcgttcggcccgacgcgttgt	962
Db	906	agtggggctgttccttggaagaccgcgcgacgctgtctgtctggcgcgaagcccgcaagctgt	965
QY	963	gttcaagaagaagacaacgctcaatgtgtcgtgcgcgcgcgcgaacccgcgaagaagcgttgcca	1022
Db	966	cgttaccacaagacagacacacatcgttcgaagggcttcgggcgatttccgattgccatgcccg	1022
QY	1023	ccgggcgaaacctgtgctgcgcgaagatcgaacaagagcgaattcggatlttgatgttggaana	1083
Db	1026	ccggggtgcttcatgtatccggcccgagatcgaaacagcgaatccgacttaccagaccgcagaa	1083
QY	1083	gcttggagagcgtgtgccaactgc	1144
Db	1086	gcttcagagagcgttgcacaaagctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	1144

OY	1143	caccgaaccgcgaactcaaggaagcgaagaaagcgtcgaaggaatggcgtgcgcgcga	1202
Db	1146	CACCGAGGTGGAGCTCAAGAGAGCGCAAGCACCGCATCGAGAGAGGCCGCTCCGCAACGGGAA	1205
OY	1203	ggccgcggtcgaaggaaggcatgcgccgtgtgtagggagcctgcgtcatccaaagggccg	1262
Db	1206	GGCTGCGCGTGGAAAGAGGGCATCTGCGCCGCGGTGGCGCGCTGTGCTGTGCAATGGCGACC	1265
OY	1263	caagagcgtcacccgaactggtgcgtgcgttgaccggtagaagaggtccctcggltgttcgaagt	1322
Db	1266	T---GCCCTGGAGCACTCGGC-----CTGACGGGGCAGCGAGGCGACCGGGGCCAAT	1316
OY	1323	gtctccgaagcccttgcgcgcgcgtgtgtctcgaagtcgcgcgaacgcttggtcttggaacg	1382
Db	1317	CGTCGCGCGTGGCGGTGTCGGCTCGCTTCAAAGCAGATCGCCTTCAMAGGGCGGCGTGGAGCC	1376
OY	1383	ctcgggtgtgtgtaacaagaatcagcgaagctaccgcgcgcggaaatggctgtgaacgtgaac	1442
Db	1377	CGGGGTGTTGCCGAGAAAGGTTGCCAACCTGCCCGGGGTCCAGGCCCTCAACAGCGCGAC	1436
OY	1443	cctgagctatgtggaacttggccgtctgaagcgtctcatcgaaccgcgcgaaggtgaatggtc	1502
Db	1437	CGGTGAGTACGAGGACTGCTCAAGGCCGCGGCTGCCGCACCGGTGAAGGTACCCCGCTC	1496
OY	1503	ggcgcgtgtgaacgcytcatcgtgttgcgcggaatgtgtaactcaacaaaggaagcgtcgtgt	1562
Db	1497	GGCGCTCAAGAACGGCGGCGTCATCGCGGCTCTGTTCTCATCACACGAGGCGCTGCTGCGC	1556
OY	1563	cgacaagccgg	1573
Db	1557	CGACAAGCCGG	1567

RESULT 10
 US-09-095-855-113
 ; Sequence 113, Application US/09095855
 ; Patent No. 6160093
 ; GENERAL INFORMATION:
 ; APPLICANT: Tan, Paul
 ; APPLICANT: Visser, Elizabeth
 ; APPLICANT: Skinner, Margot
 ; APPLICANT: Prestidge, Ross
 ; TITLE OF INVENTION: Compounds and Methods for
 ; TITLE OF INVENTION: Treatment and Diagnosis of
 ; NUMBER OF SEQUENCES: 208
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Law Offices of Ann W. Speckman
 ; STREET: 2601 Elliott Avenue, Suite 4185
 ; CITY: Seattle
 ; STATE: WA
 ; COUNTRY: USA
 ; ZIP: 98121
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: DOS
 ; SOFTWARE: FASTED for Windows Version 2.0
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/095,855
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/705,347
 ; FILING DATE: 29-AUG-1996
 ; APPLICATION NUMBER: 08/873,970
 ; FILING DATE: 12-JUN-1997
 ; APPLICATION NUMBER: 08/997,362
 ; FILING DATE: 23-DEC-1997
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Sleath, Janet
 ; REGISTRATION NUMBER: 37,007
 ; REFERENCE/DOCKET NUMBER: 11000.1002c3
 ; TELECOMMUNICATION INFORMATION:


```

;
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-09-031-606-9

```

```

Query Match      37.7%; Score 609.4; DB 3; Length 2167;
Best Local Similarity 68.4%; Pred. No. 6.9e-120;
Matches 860; Conservative 0; Mismatches 391; Indels 6; Gaps 1;

```

```

QY 4 aagctgatatgaatcgaacgaacccgctgcgcgcacatgaggtcgcacatgagcaacgctg 63
DB 855 AAGATTCTGAAGTTCGACGAGGAGCGCCGCTCGGCTTGAAGCGCGCTGAACAGCTG 914
QY 64 gccgcacacgctgcgtgtagcgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 123
DB 915 GCCGACACCGTCAAGGTTCACATCGGCCCCAAGGCGCGCACGTCCTCATCGCAAGAAG 974
QY 124 ttctgagacacacggttacacacacgagctcagcgtgcacgtgtgacatcagctgagaa 183
DB 975 TTCGGGGCCCCGACCATCATCAACAGAGGCGTACCATGCGCCGTGAGTCCAGTCCGAC 1034
QY 184 gatccgcttgaagacttggcgccagctggtgaaagtcggtgcacacacacacacacat 243
DB 1035 GACCCGTACGAGAACCTCGGCGCCAGCTGCTCAAGAGAGTGGCGACCAAGACCAACGAC 1094
QY 244 gttgcgcgtgagagcgacacacacacacacacacacacacacacacacacacacacac 303
DB 1095 ATCGCGGGGTGAGCGCACCCACCCGCGACCTGCTGCGCCAGCGCTGCTGCGCGAGGCG 1154
QY 304 ctgagcgtagtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 363
DB 1155 CTGCGGCAAGTTCGCGCGCGCGCGCTCCCGCGCGCTCAAGAGGCGATCGACGCGCGC 1214
QY 364 gccgcacgctgattcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 423
DB 1215 CTCGCGCGCGCTCGCGCGCGCGCTCGACACCGCGCGCGCGCTCGATCGACCAAGTCCGAC 1274
QY 424 atcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 483
DB 1275 ATCGCGCGCGCTCGCGCGCGCTCTCCCGCGCGAGCACAGACAGTGGCGCGCTCATCGCGCAG 1334
QY 484 gcgatgaacaaagtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 543
DB 1335 GCGATGGAACAAGTTCGAGCAAGCGTGTCAATCAAGTTCGAGAGAGTCAACACCTTCGCT 1394
QY 544 accgagttgagttacacgaaggtattgcttcacacaaagcttctgcgcgcacatctc 603
DB 1395 CTCGACCTGACTTACCGAGGCGATGCGCTTCGACAAAGGCTTACCTGCTCCCGTACATG 1454
QY 604 gttacgcacacacacacacacacacacacacacacacacacacacacacacacacacac 663
DB 1455 GTGACCGACCAAGAGCGTATGAGGCGCGCTCTCGACGACCGCTACTCTTGTATCCACGAG 1514
QY 664 gacaagatcagctgcctcccgatcgtgtgcacatgctgtaaaaggttgcgaagaaacggt 723
DB 1515 GCGAAGATGCTGTGATCCAGAGACCTGCTCGCGCTCTGAGAGAGTGTATCCAGGCGGCT 1574
QY 724 -----aagcactactgctgctgctgtaagaagctggaagcggaagcgttgcgcgcgc 777
DB 1575 GCGTCAAGCCCTGCTGTATCCCGCGAGGAGCTCGAGGCGAGGCGCTGTCGACCCG 1634
QY 778 gtcgtcaacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 837
DB 1635 GTGGTCAACAGATCCGCGCGCTTCATCAAGCGCTGCGCTCAAGGCGCGCGCTTCGCT 1694
QY 838 gacgcgcgtaagcgcttccttgaagacacgtgcgtggtgtaacggtgcgcgcgcgcgcgcgc 897
DB 1695 GACGCGCGCAAGGCGATGCTGCGGACATGCGCACCGCTCACGCGTGCATCGCGC 1754
QY 898 ccgacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 957

```

```

DB 1755 GAGAGGTGCGCGCTCAAGCTCGACCAAGCGCGCTTGGACCTGTGGACACCGCCGCGC 1814
QY 958 gttggtgtgaagaagacacacacacacacacacacacacacacacacacacacacacac 1017
DB 1815 GTACCGTCAACAAGAGCACGACGACATGTGACCTTGAAGAGAGACCGCCAGGACGTC 1874
QY 1018 gccaacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1077
DB 1875 CAGGCGCGCGTGGCCAGATCAAGCGCGAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 1934
QY 1078 gaaagcttgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1137
DB 1935 GAGAACTCTCAAGAGCGCTTCGCGCAAGCTGCGCGCGCGCTGCTGCTGCTGCTGCTG 1994
QY 1138 gccgcacacacacacacacacacacacacacacacacacacacacacacacacacac 1197
DB 1995 GCGGCGCGCGAGGTGAGTGAAGGCGCAAGCACCGCTGAGAGAGCGCCATCTCCGCG 2054
QY 1198 gccaaagccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1254
DB 2055 ACCCGCGCGCGCTGAGAGGAGGACATGCTCGGTGCTGCTCCGCGCTGCTCCAC 2111

```

```

RESULT 15
US-08-997-080-161
; Sequence 161, Application US/08997080
; Patent No. 5968524

```

```

GENERAL INFORMATION:
APPLICANT: WATSON, JAMES D.
APPLICANT: TAN, PAUL, L.J.

```

```

TITLE OF INVENTION: METHODS AND COMPOUNDS FOR THE TREATMENT OF IMMUNOLOGICALLY-
NUMBER OF SEQUENCES: 194
CORRESPONDENCE ADDRESS:
ADDRESS: Law Offices of Ann W. Spectman
STREET: 2601 Elliott Avenue, Suite 4185
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98121

```

```

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0

```

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/997,080
FILING DATE:
CLASSIFICATION:

```

```

PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:

```

```

ATTORNEY/AGENT INFORMATION:
NAME: Sleath, Janet
REGISTRATION NUMBER: 37,007
REFERENCE/DOCKET NUMBER: 11000.1007
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-269-0565
TELEFAX: 206-269-0563
TELEX:

```

```

INFORMATION FOR SEQ ID NO: 161:
SEQUENCE CHARACTERISTICS:
LENGTH: 985 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-997-080-161

```

```

Query Match      24.4%; Score 395.2; DB 2; Length 985;
Best Local Similarity 65.3%; Pred. No. 6.2e-75;
Matches 615; Conservative 0; Mismatches 316; Indels 9; Gaps 2;

```

```

QY 639 gtagcgtgtgactctgtacacacacacacacacacacacacacacacacacacacacacac 698

```

Db 1 GGATCCCTACATCCGCTGCTCAGCTCCAGGTGTCCAGCTCAAGGATCTGCTCCGCT 60
QY 659 gcttgaaaggcttcaggaagcgttaagcaactactgctgctgagcgttgagg 758
Db 61 GCTGAGAGAGTTCATCCAGGCCGAGCCCTCTCTGATCATCCGAGGACGTGAGGG 120
QY 759 cgaagcgttgcgagcgtgctgctcagcagcgttcgcaagcgttgaagcgttcggt 818
Db 121 CGAGGCTCTGTCCAGGCTGTGCTCAACAGATCCGCGACCTTCAAGTCCGTGCGCT 180
QY 819 caagggccgacttcgctgagccgctgaagcgttccttgaagaccctgagcgttgatc 878
Db 181 CAAGGCTCGGGCTTCGGTGTGACCCCGCAGAGCGATGCTGACGAGCATGCGCATCTCAC 240
QY 879 ggttggcagctggttcaaccccgagcgttcgtgctgcgagagtggttgaggt 938
Db 241 CGGTGCTCAGCTGCTCAGCAGAAAGATCGGCTGTCTCCCTGGAGACCGCGACGTCTGCT 300
QY 939 gctgggctcggcccgagcgtggtggtgagcaagagcagcagcgttcattgtcagcggc 998
Db 301 GCTGGGCGAGGCCGCAAGGCTGTGCTACCAAGAGCAGACCACTGTCGAGGGCTC 360
QY 999 cggacccgcaagacggttgcgaacccgagcgaaccacttgcgtccgagatcgaacag 1058
Db 361 GGGGATTTCCGATGCCATCGCGGCGGGGTGCTCAGATCCGCGCGAGATCGAAGACAG 420
QY 1059 cgattcggatctggatcgggaaagcgttgcgagcgttcgcaaacctgcccggcgggt 1118
Db 421 CGACTCCGACTACGACGCGCGAAGCTGCAGAGAGCGCTGGCCAAAGTGGCCGGGGTGT 480
QY 1119 tgcgtcatcaagctggtgctgcgcacccgacacgcgcaactcaagagcgaagagcgt 1178
Db 481 TGCCTGTGTCAGGCTCCGAGCTGCACCGAGAGTGCAGTCAAGGACGCAACCGCAT 540
QY 1179 cgaagatcgagctgcgagccgcaagccgagcgaagagcgttcgtccgtgctgaggg 1238
Db 541 CGAGGAGCGCGTCCGCAAGCGGAGAGGCTGCCGTGAGAGAGGCGATCGTCGGGTGGCG 600
QY 1239 agcctcgtcatcaccagagcccgcaagcgttcgcaactgcgtgcgtgcgagccgg 1298
Db 601 CGTGGCTCTGCTGCTGCTGCTC---CTGCGCTGACGACCTCG-----CTGACGGG 651
QY 1299 tgaagagctcctgctgctgagcgttctcgaagcccttgcgagcgttgcgttcggt 1358
Db 652 CGACGAGCGCCACCGGTGCGCAATGTCGCGGTGCTGCGCTGCTGCGCTCAAGCAGAT 711
QY 1359 cggcgcgaagcgtgctgagcgtgcgtgcgtggttgcgaacaggtcagcgagctaccgc 1418
Db 712 CGCCTTCAACGCGCGCTGAGAGCCCGGCGTGTGCGCAGAGAGTGTCCAACTGCCCCG 771
QY 1419 cggcgcagctgagcgttgaacacctgaactatgctgacttgcgcgttcgagcgttcac 1478
Db 772 GGGTCAACGGCGCTCAACGCGCGCGCGGTGAGTACGAGAGACTGCTCAAGCGCGGTGCG 831
QY 1479 cgaccggttcaagctgagcgttcgagcgttgcgaacgcttcacggttcgagcgttgc 1538
Db 832 CGACCCCGGTGAGGTCAACCCCTCGGGCTGCGAAGACGGCGGCTGCTCATCGCGCTGTT 891
QY 1539 actacacccagagcgttgcgtgctgcgaacagcggccaagc 1580
Db 892 CTTACCAACCGAGCGCTGCTGCGCGACACCGCGAGAGGC 933

DNA encoding a Str
 Neisseria meningit
 Neisseria meningit
 Neisseria gonorrhe
 Neisseria meningit
 Neisseria meningit
 Nucleotide sequenc
 M. vaccae antigen
 Nucleotide sequenc
 DNA encoding a Str
 Streptococcus pneu
 DNA encoding a Str
 Enterococcus faecae
 hsp60 DNA, Helico
 Hsp gene, Helicob
 H. pylori GH90 118
 Lawsonia intracell
 Lawsonia intracell
 Helicobacter pylori
 H. pylori heat sho
 Heat shock protein
 Borrelia burgdorfe
 Hsp operon, Chlam
 Hsp operon, Chlam
 M. vaccae antigen
 Nucleotide sequenc
 Complete genome s
 Heat shock protein
 Human heat shock p
 Staphylococcus aur
 Heat shock protein
 Nucleotide sequenc
 Mycoplasma genital

[illegible][illegible]


```

Db 677 gatcgccgacccgacgattcgcggtgacacatccatccgtgacatgctgcgcga 736
Qy 483 agcgtatgaacaaagtcgcccacgacgctggtcagctcaagaaatccatccagctgg 542
Db 737 ggcgtatgacaagtgaggcaagcagaggtatccacgttcgaaggtccacacatttgg 796
Qy 543 caccgattgagttccacgaaggtattgcttcacaaaggtctctgtccatctact 602
Db 797 gctgcagctcgaagctcgaaggtatgctgtccacaaaggtccatccatccgtgactc 856
Qy 603 cgttacccgacttcgaataaccagcagcggtgctcgaagacgctgtatccctacacaa 662
Db 857 cgtgacccgacccgagagtgtaagagcggtctcgtgaggaagccatccatccgtgcag 916
Qy 663 agacaagatcagctcgtctccacatctgttgcacatgctcgaagaaagttcaggaacgg 722
Db 917 ctccaaaggtgtccactgtcaagatctgctgcgcgtctcgaagaaggttcacgtgagccg 976
Qy 723 taagcactactatgactgctgctgagacgtgagagcggaagcgttggcagacgtgctgt 782
Db 977 taagcctcgtctgactcctcgcgcgaagcgtcgaagcgagcgtctgtccacccctgtgct 1036
Qy 783 caacgcgattcgcgaagcgttgaagcgttcgaagcgctcgaagggcgctactctgtgacg 842
Db 1037 caacaagatccgcgcgaccccttcaagctggtgagcggtcgaaggtctccgcttcgagccg 1096
Qy 843 ccgttaagcgtctccctgaagacgtcgcgtggtgacggtgacgaacgttgcaccccgga 902
Db 1097 ccgcaagcgatgctgacaggtatgagcattctcacccgtgtgacaggtgacggaaga 1156
Qy 903 cgcgcgactgtgctgctgcgaggtgagcgttggaggtgctgcgcgcgcgcgcgcgtggt 962
Db 1157 ggtcggcctcgaagcgtcgaagacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1216
Qy 963 ggtcgaagcaagcagacacgcgttcactgctcgaagcgcgcgacccgaagcgcgtgaccaa 1022
Db 1217 ggtcacaagagcagagacacacatcgtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1276
Qy 1023 ccggtgcgaacacactgctgctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1082
Db 1277 acgagtggtccagatccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1336
Qy 1083 gcttgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1142
Db 1337 gctgcgaagcagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1396
Qy 1143 caccgacacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1202
Db 1397 caccgaggtcgaactcaagagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1456
Qy 1203 ggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1262
Db 1457 ggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1516
Qy 1263 caagcgctcgaacgaactgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 1322
Db 1517 -----gacctggaagcagcgtgaagcgtcgaagcgtcgaagcgtcgaagcgtcgaagcgtc 1567
Qy 1323 gttctccgaagcccttgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1382
Db 1568 cgtgaaggtgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 1627
Qy 1383 ctgcgtgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 1442
Db 1628 gggcgtgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 1687
Qy 1443 ccgtgagctatgagctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 1502
Db 1688 cgtgtgtctacagagatctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 1747
Qy 1503 ggcgtgtctgaacgcgtcatcgtgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 1562

```

```

Db 1748 ggcgtcgaatgctgctgctcctgctgctgctgctgctgctgctgctgctgctgctgctg 1807
Qy 1563 cgacaagccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1591
Db 1808 cgacaagccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1836

RESULT 6
Z11371
ID Z11371 standard; DNA; 1626 BP.
XX
AC Z11371;
XX
DF 25-OCT-1999 (first entry)
XX
DE Nucleotide sequence of M. vaccae antigen GV-27.
XX
KW Mycobacterium vaccae protein; antigen; T cell activation; cytokine;
KW dendritic cell maturation; infectious disease; immune disorder; cancer;
KW respiratory system; Mycobacterium infection; allergy; tuberculosis;
KW leprosy; sarcoidosis; lung cancer; asthma; skin disorder; psoriasis;
KW dermatitis; eczema; alopecia areata; skin cancer; basal carcinoma;
KW squamous cell carcinoma; melanoma; ss.
XX
XX Mycobacterium vaccae.
OS
XX
XX WO932634-A2.
XX
PD 01-JUL-1999.
XX
PF 23-DEC-1998; 98NO-NZ00189.
XX
PR 04-DEC-1998; 98US-0205426.
PR 23-DEC-1997; 97US-0996624.
PR 23-DEC-1997; 97US-0997080.
PR 23-DEC-1997; 97US-0997362.
PR 11-JUN-1998; 98US-0095855.
PR 17-SEP-1998; 98US-0156181.
XX
PA (GENE-) GENESIS RES & DEV CORP LTD.
XX
PI Prestidige RL, Skinner MA, Tan P, Visser ES, Watson J;
XX
DR WPI: 1999-430163/36.
XX
PT P-PSDB; Y14909.
XX
XX
XX Enhancing immune response to an antigen
PS Claim 3; Page 212-213; 243pp; English.
XX
XX
XX The invention provides heat-killed Mycobacterium vaccae, or recombinant
XX M. vaccae proteins. The M. vaccae proteins may be employed to activate
XX T cells and natural killer cells, to stimulate the production of
XX cytokines, to enhance the expression of co-stimulatory molecules on
XX dendritic cells and monocytes, and to enhance dendritic cell maturation
XX and function. The proteins can be expressed by standard recombinant
XX methodology. Pharmaceutical compositions comprising the proteins or
XX nucleic acid sequences encoding the proteins can be used for the
XX treatment, prevention, and detection of disorders including infectious
XX diseases, immune disorders and cancer. In particular, the compounds and
XX methods are used for treatment of diseases of the respiratory system,
XX such as mycobacterial infections, asthma, allergies, tuberculosis,
XX leprosy, sarcoidosis and lung cancers, and disorders of the skin such as
XX psoriasis, atopic dermatitis, eczema, allergic contact dermatitis,
XX alopecia areata, and skin cancers such as basal carcinoma, squamous cell
XX carcinoma and melanoma.
XX
SO Sequence 1626 BP; 302 A; 537 C; 552 G; 235 T; 0 other;

```

```

Query Match 42.4%; Score 685.6; DB 20; Length 1626;
Best Local Similarity 65.6%; Pred. No. 1.8e-115;
Matches 1035; Conservative 0; Mismatches 534; Indels 9; Gaps 2;

```

[illegible]

OY	1083	gcttgcgaagcggcggtcccaaaccttcgacggcgagggtcttgctttcaataagtgggttcgcgc	1142
OY	1086	gcttcgaaagcgccttcggccaagcttcgaccgggtgcttgcggttgatcctaagtcggagacttcgc	1145
OY	1143	caccacacaccgcacattcaagtagagcgacaagaaaaggtctcgagatctcgggtctcgcgccgcca	1202
Db	1146	caccgaagttgtagctccaagtagagcgcaagcacccgatctgagtagacggtctccgcaacgsgaa	1205
OY	1203	ggccgctggttcgaagaaggcatctgcccttggttggtgggaagcctgcgtcataccaagagcccg	1262
Db	1206	ggcttcgcgttcgaagaaggcatctgcgcgggtgcggtcggtctctgtctgcagatcgtgc	1265
OY	1263	caaggcgtctgacccgaacctgcgttcgtctgacccgggtgacagaggtctcttcgtgtcgcgcgt	1322
Db	1266	t-----gcgctggaagcaaccttcgcgc-----ctgaacgggtgcacagagcccacgggtgcacaacat	1316
OY	1323	gtctctccaaagcccttcgacgacgcttgcttgatccgcccccaagccttgcttgagacgg	1382
Db	1317	cgctccgctgtgcgctgctgtgcgtctccgtctcaagcaaatctgcgtctcaacggcgctggagcc	1376
OY	1383	ctcggcttggttgtaacaagaagctcagcgaagctaccccgccgggtgcatctggcttgaaagtgaacac	1442
Db	1377	cggcgctgcttgacgaagaaggtgtccaaccttcgcgcgggtgtcaacgctccaacgacgacgcgc	1436
OY	1443	ccctgaagcatgtgtacttgcgcgcgttgacgggtcatctgaccgggtcgaagtgtgactggtc	1502
Db	1437	cgttgtgaagcagaagaccctgcgtcaagggcgggtcgcggatcccggttgaaagtacacccgctc	1496
OY	1503	ggcggctgttgaaacgcatcatcggttcgtcccgagatgtgtaactacacacgagacggtcgtgt	1562
Db	1487	ggcgctgcagaagcggcggtgtccatctgcgggtctgttctctcacccgagggcgtgtcgtcgc	1556
OY	1563	cgacaagccggccaaagc	1580
Db	1557	cgacaagccggagaagac	1574
RESULT	7		
V34608 ID	V34608 standard; DNA; 1569 bp.		
AC	V34608;		
DT	25-AUG-1998 (first entry)		
DE	M. vaccae antigen GV-27 encoding DNA.		
KM	Mycobacterium vaccae; antigen; therapy; prevention; cytokine production;;		
KW	M. avium; M. tuberculosis; Immune response enhancer; cell proliferation;		
KX	mycobacteria infection; vaccine; cancer; ss.		
OS	Mycobacterium vaccae.		
FH	Key	Location/Qualifiers	
FT	CDS	1..1569	
FT		/*tag= a	
FT		/product= "GV-27 antigen"	
FT		/note= "the stop codon is not indicated"	
PN	W09808542-A2.		
PD	05-MAR-1998.		
PF	28-AUG-1997;	97MO-NZ00105.	
PR	12-JUN-1997;	97US-0873970.	
PR	29-AUG-1996;	96US-0705347.	
PA	(GENE-) GENESIS RES & DEV CORP.		
P1	Hiyama J, Prestidge RL, Scott LM, Skinner MA, Tan P,		
E:	Visser E;		
X			

QY	1323	gcttcgcgaagcccttgcgcgcgcgtgtgtcttgatcgcgcgcgaacgcctcgtcttggacgg	1382
QY	1323	gcttcgcgaagcccttgcgcgcgcgtgtgtcttgatcgcgcgcgaacgcctcgtcttggacgg	1382
Db	1317	cgctccgcgtgcgcctgtcgcctcgcctcgaagcagactccttcaacggcgccttggagcc	1376
QY	1383	ctcgcgtgtgttgcacacaaggttcagcgagctacccgcgcgcagatctggaactgaacac	1442
Db	1377	cgagcgtcgttgc	1436
QY	1443	cctgaagctatgttgaacttgcgcgccttgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	1502
Db	1437	cggtgagctacagagaccctgtcctcaagcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	1496
QY	1503	ggcgcgttgaacgc	1556
QY	1563	cgacaagccgcg	1573
QY	1567	cgacaagccgcg	1567
RESULT	9		
ID	N80339	standard; DNA; 3613 BP.	
XX	N80339;		
XX	10-OCT-1990	(first entry)	
DT	Clone Y3178	insert contg. M. leprae 65kD antigen gene.	
DE	Leprosy; antigen; vaccine; armadillo; ss.		
XX	Mycobacterium leprae.		
OS			
XX	Key	Location/Qualifiers	
FT	CDS	66..1832	
FT		/*lag= a	
FT	misc_RNA	/product=61,856 dalton protein	
FT		207..1832	
FT		/*lag= b	
FT		/product=56,686 dalton protein	
PN	MOB800974 -A.		
XX	11-FEB-1989.		
PD			
XX	28-JUL-1987;	87MO-US01825.	
XX			
PR	31-JUL-1986;	86US-0892095.	
XX			
PA	(WHIT-) WHITEHEAD INST BIOM.		
XX			
PI	Young RA;		
XX			
DR	WPI: 1988-049978/07.		
DR	P-PSDB: P80364.		
XX			
PT	DNA encoding antigens of Mycobacterium leprae - used for producing		
PT	immuno-determinant protein antigen for prevention, diagnosis and		
PT	treatment of leprosy.		
PS	Claim 7; Fig 1; 46pp; English.		
XX			
CC	The gene was isolated by probing a lambda gt11 library of M. leprae		
CC	DNA with Mabs directed against M. leprae specific antigens. The		
CC	sequences were determined with a rapid primer extension method.		
CC	The antigen appears as a doublet on SDS PAGE with mol. wt. 55-65 KD		
CC	so translation may initiate at the start of either CDS in the feature		
CC	table, producing two polypeptides. The sequence can be used in the		
CC	development of highly specific serological tests for screening		
CC	populations for individuals producing Abs to M. leprae, in the		

CC development of vaccines and in the assessment of efficacy of treatment.
XX
SQ Sequence 3613 BP; 724 A; 960 C; 1183 G; 746 T; 0 other;

Query Match 39.28; Score 634.4; DB 9; Length 3613;

Best Local Similarity 63.48; Pred. No. 3e-106;

Matches 1006; Conservative 0; Mismatches 571; Indels 9; Gaps 2;

[illegible]


```

Db 1695 gaccgcgcaagcgatgctgcgcacatgagccacccctacacggtgtccacggtatcgcgc 1754
QY 898 ccgacgcgcgcatgtgctgcgcgaggtggtgcttgaggtgtcgtggtcgcgcgcgcgc 957
Db 1755 gaggaaggtcggctcctcaagctcagaccagcgctgtgacgtgtcgtggcaacgcccgcgc 1814
QY 958 gtgtgttcacgaaggaagcagacgctatctgtcagcgcgcgcgcgcgcgcgcgcgcgcgcgc 1017
Db 1815 gtacccgtcacaagaaggaagcagacacatcgtgagcgcgcgcgcgcgcgcgcgcgcgcgcgc 1874
QY 1018 gccacacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1077
Db 1875 caggcgccgctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1934
QY 1078 gaaagctgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1137
Db 1935 gagaaagctcacaaggaagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1994
QY 1138 gccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1197
Db 1995 gcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 2054
QY 1198 gccaaagcgcggtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1254
Db 2055 acccgcgcgcggtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 2111

RESULT 11
Q22481
ID Q22481 standard; DNA; 1320 BP.
XX
AC Q22481;
XX
DT 29-JUL-1992 (first entry)
XX
DE groEL-1 gene partial sequence.
XX
KM Heat shock protein; HSP18; ss.
XX
OS Streptomyces albus.
XX
FH Key Location/Qualifiers
FT CDS 1..510
FT /tag= a
FT /note= "groEL-1 gene encoding HSP18"
XX
GN MO9204452-A.
XX
NC 19-MAR-1992.
XX
PE 03-SEP-1991; 91MO-FR00701.
XX
PR 10-SEP-1990; 90FR-0011186.
XX
PA (INSP ) INST PASTEUR.
XX
PI Mazodier P, Guglielmi G;
XX
DR WPI; 1992-114358/14.
XX
DR PSDB; R22362.
XX
PT Recombinant DNA contg. heat inducible promoter and heterologous
PT gene - also vectors, transformed cells and new heat shock
PT proteins of Streptococcus albus
XX
PS Claim 18; Fig 6; 50pp; French.
XX

```

This is a partial sequence of the groEL-1 gene which codes for the 18 kDa heat shock protein HSP18. The gene actually codes for a 56 kDa protein but this is subjected to post translation modification to give the 18 kDa HSP18 protein, the partial sequence given here is missing the last 300 nucleotides encoding the C-terminal of the

```

CC 56 kDa protein. See also Q22477-Q22486.
XX
SQ Sequence 1320 BP; 237 A; 469 C; 444 G; 170 T; 0 other;

Query Match      38.2%; Score 617.4; DB 13; Length 1320;
Best Local Similarity 68.8%; Pred. No. 3.3e-103;
Matches 865; Conservative 0; Mismatches 386; Indels 6; Gaps 1;

QY 4 aagctgtacgaatacgaagaacccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 63
Db 7 aagatctgaagttcagcaggaagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 66
QY 64 gccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 123
Db 67 gccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 126
QY 124 ttggcgagaccacacgcgttaccacgaagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 183
Db 127 ttgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 186
QY 184 gatccgtttgaagaacttgggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 243
Db 187 gaccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 246
QY 244 gtggcggtacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 303
Db 247 atcggggtacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 306
QY 304 ctgaggtacgttggcgcgcgcggtaccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 363
Db 307 ctggcgcaacgctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 366
QY 364 gccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 423
Db 367 gtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 426
QY 424 atcgcgcaagttgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 483
Db 427 atcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcg 486
QY 484 gccgtgacgaaggtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 543
Db 487 gcgatgacgaaggtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 546
QY 544 accgagttgaggttcacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 603
Db 547 gtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 606
QY 604 gtacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 663
Db 607 gtgacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 666
QY 664 gacaaagatcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 723
Db 667 ggcgaagatcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 726
QY 724 -----aagcactacgtatcgttgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 777
Db 727 ggtcccaagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 786
QY 778 gtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 837
Db 787 gtgtgtcacaagaagatcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 846
QY 838 gccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 897
Db 847 gaccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 906
QY 898 ccgacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 957
Db 907 gaggaaggtcggctcctcaagctcagaccagcgctgtcgtgagcgtgtcgtggcgcgcgcgcgcgcgcgcgc 966

```



```

QY 434 tggcagcgtgtctc---gcgcagcagcagatcgttgaacctgtgtgscgaacgata 490
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 454 ttgttaccatctccgtaactcgaacgaacgtagttaactgtatcgtcgtgaacgata 513
QY 491 acaagtcgcgcacacagcagctgtgttcagcgtlcaagatccctccacgcgtgscacgagt 550
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 514 acaagtcgcgtlaagaaagcgtgtatccacgcttgaagacggtlaccggtctgcaagcgaac 573
QY 551 tggatgtaccacgaagatattgtgtcccaaaaggtctctgtcggaactcgttaccg 610
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 574 tggacgtgtgttgaaggtatcagttcgaacgtgtgttaccctgtcttaccatccaca 633
QY 611 acctcgaataacacgaagcgtgtcgaagcgtgtgtatcgtgttaccacacgaacaga 670
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 634 agccggaacgtgcgcagcagtagaattggaagcccggttccatcccgctggtcgaacaga 693
QY 671 tcaagtcgtctcccgatctgttgcacatgtctggaanaaggttgcaggaacggttaacac 730
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 694 tctcaacatccgcgaatagtctgcggttctgtggaagctgtgagaaagcaggaacacgc 753
QY 731 tactgtatcgtgtgtgaagacgttgaagcgaagcgttgcgcagcgtgtgtcgtcaacgca 790
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 754 tgtgtatcatcgtcgtgaagatgttgaagcgaagcgtgcgaactctgttcttaccacca 813
QY 791 ttgcgaagcgttgaagcgttcgcgcgtlcaagggcgttactcgtgtgacccgcgttaag 850
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 814 tgcgcgtgtatcgttaaaagtcgtcgtgttaagacactgtcgtccgcgtcgtcgttaag 873
QY 851 cgttctcttgaagacgttgcgcgtgtgtgaagcgtgcgcagcgtgttcaacccgcagcga 910
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 874 caatgtcgcagagatcgtatcgaacctgtgcgcgtgtgttaccgttactctcgaagagatcgt 933
QY 911 ttgtgtcgtcgcagagcgtgtgaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 970
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 934 tggagctgtgaaagacactctgtgaagatctgtgcccagcgtgtgtgttatacaca 993
QY 971 aggaacgacagcgtatctgtcgaagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1030
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 994 aagatataccacacatcatcgaatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1053
QY 1031 accactgtcgttgcgcagatcgcacgaagcgtatcgaattgtgaltcgggnaaagctgtgc 1090
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1054 ctacagatctgtcagacagatcgaagaagcaacttccgactatgaccgtgaanaactgtgag 1113
QY 1091 agcgtcgtcgcacaaatgtgcgcgcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1150
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1114 agcgtcgttgaagaaatgtgcgcgcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1173
QY 1151 ccgcactcaaggaagcgcgaagaaagctgcgaatgtgcgcgcgcgcgcgcgcgcgcgcgcgc 1210
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1174 ttgaatgaaagaaagaaagccgcgttgaagatgtccctgcacgcgttaccgtgtcgcgt 1233
QY 1211 tcgaagagggcgtatcgttccctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1270
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1234 tagaagaagcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1287
QY 1271 tggacgcagcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1330
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1288 aaattgcgcgcgtgaaggtlcaagaacgaacgaacgaacgaacgaacgaacgaacgaacga 1347
QY 1331 aaacccctgtgcgcgcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1390
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1348 ggcgaatggaatcccccctgtgtcaaatcgttactgaactcgtgcgcgaagagcgtctgtag 1407
QY 1391 tggatcaacaaggtcagcgtatccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1450
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1408 tggctaacacacgtgaaagcgcgtgtacgttacaacgtgtcaacgtgtcaacgtgaagaa 1467
QY 1451 atgtgtactgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1510
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1468 acggaacatgatacgtatcgtgtatctcgtgtatccaaacgaacgaacgaacgaacgaac 1527
QY 1511 tgaagcgtcatcgtgtgtccgcgtgtgttactcaaccacgcgaacggtgtgtgtgacaagc 1570

```

```

Db 1528 agtaacgcgtcttctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1587
QY 1571 cg 1572
  ||
Db 1588 cg 1589

```

```

RESULT 15
254509
ID 254509 standard; DNA; 1635 bp.
XX
AC 254509;
DT 21-MAR-2000 (first entry)
XX
DE Neisseria meningitidis ORF 982 partial DNA sequence SEQ ID NO:2965.
XX
KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW antibiotic; diagnosis; immunogenic; infection; meningitis; septicemia;
KW antibacterial; gene therapy; ds.
XX
OS Neisseria meningitidis.
XX
PN W09957280-A2.
XX
PD 11-NOV-1999.
XX
PE 30-APR-1999; 99NO-US09346.
XX
PR 01-MAY-1998; 98US-0083758.
PR 31-JUL-1998; 98US-0094869.
PR 02-SEP-1998; 98US-0098994.
PR 02-SEP-1998; 98US-0099062.
PR 09-OCT-1998; 98US-0103749.
PR 09-OCT-1998; 98US-0103794.
PR 09-OCT-1998; 98US-0103796.
PR 25-FEB-1999; 99US-0121528.
XX
PA (CHIR ) CHIRON CORP.
PA (GENO-) INST GENOMIC RES.
XX
PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
PI Petersen J, Pizze M, Rappuoli R, Ratti G, Scalzo E, Scarselli M;
PI Tettelin H, Venter JC;
XX
DR MPI: 2000-062150/05.
DR P-PDB: Y75747.
XX
PT Novel Neisserial polypeptides predicted to be useful antigens for
XX vaccines and diagnostics -
XX
PS Claim 7: Page 1387-1388; 1453pp; English.
XX
CC 253015 to 254536, 254577 to 254615, and Y74253 to Y75941 represent
CC novel Neisseria meningitidis and N. gonorrhoeae polynucleotides and
CC polypeptides. 254537 to 254576 and 254616 to 25473 represent PCR
CC primers used in the exemplification of the present invention. The
CC polypeptides, the polynucleotides, antibodies and compositions of
CC the invention can be used as vaccines, as diagnostic reagents, and as
CC immunogenic compositions. The polypeptides can be used in the
CC manufacture of medicaments for treating or preventing infection due to
CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the
CC presence of Neisseria bacteria, or to raise antibodies. They may also
CC be used to screen for agonists or antagonists, which may themselves
CC have use as antibacterial agents. The polynucleotides of the invention
CC may also be used in gene therapy protocols.
XX
SQ Sequence 1635 bp; 448 A; 440 C; 456 G; 291 T; 0 other;

```

Query Match 29.1%; Score 470.2; DB 21; Length 1635;
 Best Local Similarity 57.6%; Pred. No. 1.1e-76;

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd

Om nucleic - nucleic search, using sw model

Run on: January 12, 2001, 12:43:46 ; Search time 6876.06 Seconds
(without alignments)
1203.509 Million cell updates/sec

```
Title: US-09-461-774-1
Perfect score: 1617
Sequence: 1 agcaagctgatacgatacga.....accacgycgacgcgcactga 1617
```

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1118133 seqs, 2558875100 residues

total number of hits satisfying chosen parameters: 2236266

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

1:	gb_ba1.*
2:	gb_ba2.*
3:	gb_om.*
4:	gb_ov.*
5:	gb_ph.*
6:	gb_pi1.*
7:	gb_pi2.*
8:	gb_pi1.*
9:	gb_pi2.*
10:	gb_pi2.*
11:	gb_pi3.*
12:	gb_sy.*
13:	gb_un.*
14:	em_fun.*
15:	em_hum1.*
16:	em_hum2.*
17:	em_in.*
18:	em_om.*
19:	em_ov.*
20:	em_ov.*
21:	em_pst.*
22:	em_pi.*
23:	em_pi.*
24:	em_to.*
25:	em_scs.*
26:	em_sy.*
27:	em_un.*
28:	em_un.*
29:	gb_hg1.*
30:	gb_hg2.*
31:	gb_hl1.*
32:	gb_hl2.*
33:	em_ba1.*
34:	em_ba2.*
35:	em_hum3.*
36:	em_hum4.*
37:	gb_py4.*
38:	gb_hg3.*
39:	gb_hg4.*
40:	gb_hg5.*
41:	gb_hg6.*
42:	gb_hg7.*
43:	em_hg1.*

45:	em_htrg2.*
44:	em_htrg3.*
43:	em_htrg4.*
42:	em_htrg5.*
41:	em_htrg6.*
40:	em_htrg7.*
39:	em_htrg8.*
38:	em_htrg9.*
37:	em_htrg10.*
36:	em_htrg11.*
35:	em_htrg12.*
34:	em_htrg13.*
33:	em_htrg14.*
32:	em_htrg15.*
31:	em_htrg16.*
30:	em_htrg17.*
29:	em_htrg18.*
28:	em_htrg19.*
27:	em_htrg20.*
26:	em_htrg21.*
25:	em_htrg22.*
24:	em_htrg23.*
23:	em_htrg24.*
22:	em_htrg25.*
21:	em_htrg26.*
20:	em_htrg27.*
19:	em_htrg28.*
18:	em_htrg29.*
17:	em_htrg30.*
16:	em_htrg31.*
15:	em_htrg32.*
14:	em_htrg33.*
13:	em_htrg34.*
12:	em_htrg35.*
11:	em_htrg36.*
10:	em_htrg37.*
9:	em_htrg38.*
8:	em_htrg39.*
7:	em_htrg40.*
6:	em_htrg41.*
5:	em_htrg42.*
4:	em_htrg43.*
3:	em_htrg44.*
2:	em_htrg45.*
1:	em_htrg46.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result	No.	Score	Query	Match	Length	DB	ID	Description
c	1	1589.8	98.3	33818	2	MTEY78	Z77165	Mycobacteri
	2	1588.2	98.2	2987	2	MTEGROEP	X60350	M. tuberculosis
	3	1074.2	66.4	2912	2	MIGROELD	Z11665	Mycobacteri
c	4	1074.2	66.4	36947	73	U00020	U00020	Mycobacteri
	5	1074.2	66.4	42325	73	U00015	U00015	Mycobacteri
	6	741.8	45.9	2286	73	SMGROELX	M76557	Streptomyces
c	7	741.6	45.9	1620	81	AR028760	AR028760	Sequence
	8	727.4	45.0	2668	81	AR028761	AR028761	Sequence
	9	724.4	44.8	41055	73	SC6G4	AL031317	Streptomyces
c	10	709.8	43.9	2431	2	MSGBCG	M17705	M. bovis BCG
	11	709.8	43.9	4380	2	MSGCMQPA	M15467	M. tuberculosis
	12	709.8	43.9	4380	81	108847	M15467	M. tuberculosis
c	13	709.8	43.9	4380	81	108856	108856	Sequence 5
	14	709.8	43.9	19781	2	MTV037	AL021932	Mycobacteri
	15	708.2	43.8	3000	73	SCGROEL	X75206	S.coelicoides
c	16	708.4	43.8	1623	81	AA6454	AA6454	Sequence 1
	17	704.6	43.6	1624	2	AP281650	AP281650	Mycobacte
	18	704.6	43.5	45313	73	SGDP95A	AL357432	Streptomyces
c	19	698.8	43.2	42741	2	MSG423	ABD00014	Mycobacteri
	20	698.2	43.2	1751	2	MPHS95	X74518	M.paratuberculosis
	21	688.6	42.6	2641	2	MPU15989	U15989	Mycobacteri

22	688	42.5	1814	73	STMGR0ELX	M7658 Streptomyc
23	685.6	42.4	1626	81	AR080045	AR080045 Sequence
24	685.6	42.4	1626	81	AR085971	AR085971 Sequence
25	681.8	42.2	1569	81	AR080017	AR080017 Sequence
26	681.8	42.2	1569	81	AR085943	AR085943 Sequence
27	680.6	42.1	2440	73	SLGROESL	X55970 S. lividans
28	674.6	41.7	2192	73	SLGROELG	X55971 S. lividans
29	665.8	41.2	1620	73	TTU90204	U90204 Tsukamurell
30	634.4	39.2	3613	2	MSGANM	M14341 M. leprae 65
31	634.4	39.2	38065	2	MICB1450	AL035159 Mycobacte
32	627.4	38.8	1688	1	AB039274	AB039274 Propionib
33	623.8	38.6	1320	81	AR028758	AR028758 Sequence
34	609.4	37.7	2167	81	AR028759	AR028759 Sequence
35	607.2	37.6	1740	81	I01709	I01709 Sequence 4
36	605.8	37.5	1950	73	TTH250409	AJ250409 Thermus t
37	604.2	37.4	2308	73	TTTHCA	D45880 Thermus the
38	601.6	37.2	1952	1	AP222061	AP222061 Propionib
39	601.6	37.2	2824	1	AB036414	AB036414 Propionib
40	597.8	37.0	2322	73	TAU29483	U29483 Thermus aqu
41	593	36.7	2769	1	AF145252	AF145252 Rhodother
42	586.8	36.3	2514	1	AB015985	AB015985 Paracoccu
43	572	35.4	2253	2	CYNGROESLA	M99443 GROEST. Chr
44	568.8	35.2	2280	2	BP012277	U12277 Bordetella
45	561.8	34.7	11884	1	AE004854	AE004854 Pseudomon

ALIGNMENTS

RESULT 1
LOCUS MTCY78 33818 bp DNA BCT 17-JUN-1998
DEFINITION Mycobacterium tuberculosis H37Rv complete genome; segment 145/162.
ACCESSION Z77165 AL123456
VERSION Z77165.1 GI:3261609

KEYWORDS
SOURCE
ORGANISM
Mycobacterium tuberculosis.
Mycobacterium tuberculosis
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacterineae; Mycobacteriaceae;

REFERENCE
AUTHORS
1 (bases 1 to 33818)
Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C.,
Harris, D., Gordon, S.V., Eigmeier, K., Gas, S., Barry III, C.E.,
Tejeda, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T.,
Conor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S.,
Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J.,
Moule, S., Murphy, L., Oliver, S., Osborne, J., Quail, M.A.,
Rajandream, M.A., Rogers, J., Rutter, S., Seeger, K., Skelton, S.,
Squares, S., Squares, R., Sulston, J.E., Taylor, K., Whitehead, S. and
Barrell, B.G.
Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence
Nature 393 (6685), 537-544 (1998)

TITLE
JOURNAL
MEDLINE
REMARK
Erratum: [[published erratum appears in Nature 1998 Nov
12;396(6707):1901]]
2 (bases 1 to 33818)
Parkhill, J.

REFERENCE
AUTHORS
JOURNAL
MEDLINE
REMARK
Direct Submission
Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium
tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome
Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique
Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux,
75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk
On Jun 27, 1998 this sequence version replaced g1:1449358.

COMMENT
Notes:
Details of M. tuberculosis sequencing at the Sanger Centre are
available on the World Wide Web.
(URL: http://www.sanger.ac.uk/Projects/M_tuberculosis/) CDS have
been renumbered from the original cosmid submissions but the old
gene designations are in brackets after the new gene numbers.
Gene prediction was based on a Hidden Markov Model of TB genes

FEATURES

implemented in TParse (Krogh) supplemented with visual inspection
of positional base preference in codons, especially where there is
an increase in the observed/expected third position G + C.
CAUTION: In some cases we may not have predicted the correct
initiation codon. Where possible we choose an initiation codon
(atg, gtg, or tlg) which is preceded by an upstream ribosome
binding site sequence (optimally 5-13bp before the initiation
codon). If this cannot be identified we choose the most upstream
initiation codon.

Location/Qualifiers
1..33818

/organism="Mycobacterium tuberculosis"
/strain="H37Rv"
/db_xref="taxon:1773"
/clone="Y78"
complement(147..1724)
/gene="guaA"
complement(147..1724)
/gene="guaA"
complement(147..1724)
/note="RV3396c, (MTCY78.32), len: 525, guaA, gmp
synthase, highly similar to Mycobacterium leprae U00015_15
GUA, MYCLE P46810 gmp synthase (glutamine-hydrolyzing)
(590 aa) Fasta scores, opt: 3000, z-score: 3351.4, E(): 0.
(86.3% identity in 531 aa overlap); contains P500442
glutamine amidotransferases class-I active site"

/codon_start=1
/transl_table=1
/product="guaA"
/protein_id="CAB01027.1"
/db_xref="GI:1449391"
/db_xref="SWISS-PROT:O50729"
/translation="MVOPADIDVETPEPARPVLVDFGQVYQALIAKIVREARVSEVI
PHTASIEIRAROPVALVLSGGPASPVAAGAPKIDPALDLGVPVLGICGFQMAOA
LGIVAFHRCRGREYRTKLVYGLKHSIDLPVQVWMSHGCAVPAADGPPVVASG
APVAFEFDRRLGVOVTHPEVHTPHQOVLAFRLHDFAGLGQVTPANTANALIEQ
VRIQIGDHAIICGLSGGVDSVAVALVQRAIGRLTCVVDHGLLRGERAQVORDV
AATGANLVYDAIETFLFELSAGVAPGKRIIRGFRAFGAVRVDLQKTEFLV
OGTLVPDVESGGSGGATNIRKSHNVGGLPDLEFTYVPELRLFKFQVAVAGREL
PEIIVARQPPGPGIGIRIVGEVTAKRIDLRIHNSDIYRELTAAIGDINOIQCQPVV
LADRVSYGVGDGRTYGHPIYLRVYSSDANTADMTVPYEVLERISIRITNEVAEYN
RVLDITSKPPATIEWE"

complement(1428..1463)
/gene="guaA"
/note="P500442 Glutamine amidotransferases class-I active
site"

complement(1736..2644)
/gene="phya"
complement(1736..2644)
/gene="phya"
complement(1736..2644)
/note="RV3397c, (MTCY78.31), len: 302, phytoene synthase,
similar to many, e.g. PSY2, LYCES P37273 phytoene synthase
2precursor (310 aa) Fasta scores, opt: 421, z-score:
500.0, E(): 6.6e-21, (32.9% identity in 295 aa overlap);
contains P501045 Squalene and phytoene synthases signature
2"

/codon_start=1
/transl_table=1
/product="phya"
/protein_id="CAB01026.1"
/db_xref="GI:1449390"
/db_xref="SWISS-PROT:O50728"
/translation="MTEIDQVRYTESITRTAARNFYGIIRLLPREKRAALSAVALG
RRIDVDAGELAPETKITELDAIKRSDIDSSDPLVAVLADAAARPPVIAFAEL
IDGARMEIDWPCGDPDELIVYCRGAGTCTICTGEGVSTASVARDIGITALQO
TNIIRVDREDFLNKRITVLRDELDRLGVRLRUDDTGLDPPDGLALILRSADRAD
WYSGILRIPIHLDRRSACCAAMSGITRRDLALIRASPAVYVDRISLSGLKQOVA
AALASVTCGPAGHPLPADLGSHPSH"

complement(2108..2185)
/gene="phya"
/note="P501045 Squalene and phytoene synthases signature
2"
complement(2673..3752)
/gene="idsA"

misc-feature
gene
CDS

Qy	841	cgcgtaagagcggttcctcttgagagacctgcyggtggtgagcgggtggtgcaacgtgtgtcacaacccc	900
Qy	23694	CGCCGTAAAGGGGTTTCCTTGAAGACCTTGCGCTGGTGTGACGGGGTGGCCAGGTGGTCAACACCC	23635
Qy	901	gaacccgcacatgtgcgtgcgcgaagtgaggtcttggaagtgcttgagctcgcgcgcgaacgctg	960
Db	23634	GACCCGGGCAATGTGTCTGTCCGAGAGTGGCTTGGAAGTGTCTGGGCTTCGGCCGACCGCTG	23575
Qy	961	gtgttcagcaagagacgacacagctcatgtctcagcggcggcgacacgcgcagaagcgtgtgc	1020
Db	23574	GTGGTCAAGCAAGAGACGACACGGTCATTGTGCAGCGCGCGGACCCGACAGAAAGCGGTGGCC	23515
Qy	1021	aaccggcgacacacacactgtcgttgcgcgagatcgcacaagaagcgtatcggattggtatcgagaa	1080
Db	23514	AACCGGGGGAAGCACTTTCGCTGGCCGGAATGCACAAAGCCATTTCGGATTGGATTCGGGAA	23455
Qy	1081	aagcttgcgagcgagcgctgtgcgaacactgcccgcggcgaggtgtcttcaataagrtgggtgc	1140
Db	23454	AAGCTTGCGAGCGGGCTGGCCAAACTGGCCGGCGGGGTCTCTCATCAAGGTGGGTGCC	23395
	1141	gccacgcacacacgcacacacacacgaagcgcgaagaaagcgtcgaagatgcygtcgcgcgcgc	1200
Db	23394	GCCACCGAGACCGCAGCTCAAGAGACGCGAAGAAAGCGTCAGAGATCGGTGGCGCGCGCC	23335
Qy	1201	aagcgcgcgatcgagagagagcatcgttcctcgtgtggggagcctcgtcatcacaagcc	1260
Db	23334	AAGCGCGCGTCTGAGAGAGGCGATCGTCCGTGGGGGAGCCCTCGCTCATCCACAGGCGC	23275
Qy	1261	cgcgaagcgctgtaccgaacatgctgtgctgtacacgggtgacgaaggttcctcgtgtgcac	1320
Db	23274	CGCAAGGCGTGTGACCGAAGCTGCTGCTGCTGACCGGTGACGAGGTCTCTGGTGTGAGAC	23215
Qy	1321	gtgtcttcgcgaagccctctgcgcgcgcgtgttctctgatacgcgcgcacacgcttgcttgagc	1380
Db	23214	GTGTTCCTCCGAAGGCCCTTGCCGCGCGCGCTGTGTGATGCCGCCCAAGCGTGGCTTGGAC	23155
Qy	1381	ggtcctggtgtgtgtcaacaaggtcagcgaactacacgcgcggcgatggtgctgaaagtgaac	1440
Db	23154	GGCTCGGTGGTGGTCAACAAGGTGACGAGACTACCCCGGGCATGTGGCTGAACCTGTAAC	23095
Qy	1441	aaccttgagcctatgttgacttggccgcctgcgaagcgctacatgcacccggttaagrtgataag	1500
Db	23094	ACCCGTGACTATGGTGTACTTGGCCGCTGACGCGCTCATTCACCCCGGTCAAGGTGACTAAG	23035
Qy	1501	tcggcggtgtgtgaacgcgtcatcgtgttcgcgcgagatggttactcacacgcagagacgtcgtg	1560
Db	23034	TCGGCGGTGTGAACGGGTGCTATGGTGGCCGGATGTGACTATCACACCGAGAGGTGTG	22975
	1561	gtcgacaagcgcgcgaagagcagaagatcacgacatcatcacacgcggcagcgacgtcgtgaa	1617
Db	22974	GTCGACAAGCGCGGCAAGAGCAGAGATCACGACCATCATCACCGGACGCGGACACTGA	22918
RESULT	2		
MTGROEOP			
LOCUS	MTGROEOP	2987 bp	DNA
DEFINITION	M.tuberculosis groE gene for KCS and 10-kDa products.	BCT	23-APR-1992
ACCESSION	X60350		
VERSION	X60350.1	GI:44599	
KEYWORDS	chapteronlin; groE gene; groESL operon; heat shock protein.		
SOURCE	Mycobacterium tuberculosis.		
ORGANISM	Mycobacterium tuberculosis		
	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;		
	Actinomycetales; Corynebacteriineae; Mycobacteriaceae;		
	Mycobacterium.		
REFERENCE	1 (bases 1 to 2987)		
AUTHORS	Shinnick,T.M.		
TITLE	Direct Submission		
JOURNAL	Submitted (26-SEP-1991) T.M. Shinnick, Center for Disease Control		
	Mailstop G35, Div of Bacterial and Mycotic Diseases, 1600 Clifton		
	Road, N.E. Atlanta, GA 30333, USA		
REFERENCE	2 (bases 1 to 2987)		
AUTHORS	Kong,T.H., Coates,A.R.M., Hickman,C.J. and Shinnick,T.M.		

FEATURES	source	Location/Qualifiers
TITLE	The Mycobacterium tuberculosis groE operon	
JOURNAL	Unpublished	
FEATURES		
source		1..2987
		/organism="Mycobacterium tuberculosis"
		/strain="Erdman"
		/db_xref="taxon:1773"
		/cell_line="E.coli XL-1"
		/clone_lib="gt11: M. tuberculosis"
		/clone="pRL4, lambda SK34"
		143..150
		162..175
		340..344
		354..656
		/codon_start=1
		/transl_table=1
		/product="10-kDa antigen homologue"
		/protein_id="CAA42908.1"
		/db_xref="GI:581363"
		/db_xref="SWISS-PROT:P09621"
		/translation="MAKVNPKPLEDKILLVQANEAFYTAGSLVLPDYAKERPOEGTVVAVGGRWDEDEDEKRIPLDVAEGDVIYKSGTEIKYNGEYILLISARDVLAVYK"
		740..744
		751..2370
		/codon_start=1
		/transl_table=1
		/product="KCS"
		/protein_id="CAA42909.1"
		/db_xref="GI:44601"
		/db_xref="SWISS-PROT:Q59573"
		/translation="MSKLIETDYARRAMEYMDKLADYRVYTLGPRGHVYLAKANGGPNVADGVVAREILEDEPEDIGADLVSVARRNDVADGTTATYTLAOLIKGGVLELAAGVNPITALGVGIKADAVSEALASATPVSKTGIAQAVTSSRDEIGDLVLGEASKYGHGVSVSEESTLGTLEFTEGIGDFKGLSAIVTFDPOQAVLEDAALI LHDOKTISLPDLPLLEKVAATGCKPLLIYAEVDEGALATLVNATIRKTLKAAVAVG PYRDEKARPLEDIAVYTGVOVYPRGMDKQVREGLVIGSARVYVSKDQTVVDGGGCTAAVYANRAKHLAEITDKSDSDMDRKLGERLAKTLGGAIVITVGAATEALKERKE SVEDEAAAKAAVEEGIVPVGGSALILQARKALTELASLTGDEVLCVDVFESLAP LFTVANAAGLDGSSVYVKVSELPAGHGLINVTYLSYGLADGVIDPVKTVSAVLNMS SVARMVLTTEYVVVDPAKAKEDHDDHHGHGH"
		2394..2419
terminator		548 a 892 c 1048 g 499 t
BASE COUNT		548 a 892 c 1048 g 499 t
ORIGIN		
Query Match	98.2%	Score 1588.2; DB 2; Length 2987;
Best Local Similarity	98.9%	Pred. No. 6.6e-168;
Matches 1599; Conservative	0; Mismatches 18; Indels 0; Gaps 0;	
OY	1	agcaagctgatacgaatacagcaagaacccgcgcgcgcatactggaagtcgcgatacgaacaag 60
DB	754	AGCAAGCTGATCGAATACGACGAACCCGCGCGTCGCATGAGAGTCGCGCATGACAAG 813
OY	61	ctgaccgacacccctgcggtgtgaacgtctggggccgcgcgcgcgcgacatgtgtctgtgccaag 120
DB	814	CTGGCCGACACCCGTCGGGTGACGCTGGGGCCGCGCCGCGCATGTGCTGGCCAAAG 873
OY	121	gcgatttcgcgagcccaacgcggtatacgaagagcgctcaacgctgtgagacgctgagatacagctg 180
DB	874	GCGTTTGGCGGACCCACGCTTACCAACACGCGCTACAGGTGGCCACGTGAGATCGAGCTG 933
OY	181	gaagaatccgatttgaagaacttggcgcccacagcttgttgaagtcggttggccaacaagaaccaac 240
DB	934	GAAGATCCGTTTGAAGACTTGGGGCCCAAGCTGTGGAAGTGGTGGCCACCAAGACCAAC 993
OY	241	gattgttcgcggtgacgcgacccaacacgcgcaacacatcttggcgcaaggaactatcaagaagc 300
DB	994	GATGTGCGCGGTGACGCGACCAACCCCAACCACTTGTGGCGCAGGCACTGATCAAGGCG 1053
OY	301	ggcctggaagctagtgagcgcgcgcgttaaacccgcgtgcgcgtccgcgttgaggaaatcggaagc 360
DB	1054	GGCCTGAGGCTAGTGTGGCCGCGGCGCTCAACCCGATGCGGCTCGGCTGGGGAATCGGCAAG 1113

[illegible][illegible]

TITLE	Direct Submission
JOURNAL	Submitted (01-NOV-1993) Department of Genetics, Harvard Medical School, 200 Longwood Avenue, Boston MA 02115
REFERENCE	4 (bases 1 to 36947)
AUTHORS	Robison, K.
TITLE	Direct Submission
JOURNAL	Submitted (01-MAR-1994) Department of Genetics, Harvard Medical School, 200 Longwood Avenue, Boston MA 02115
COMMENT	On Mar 31, 1994 this sequence version replaced gi:414226. This sequence data was produced by the Genome Sequencing Center located at Collaborative Research Incorporated (1365 Main St., Waltham MA, 02159). 617-487-7979). Please contact Doug Smith (smith@cr.cric.com). The annotation should be considered preliminary and incomplete.
FEATURES	Location/Qualifiers
Source	1..36947
	/organism="Mycobacterium leprae"
	/db_xref="taxon:1769"
	complement(1..791)
CDS	/codon_start=-1
	/transl_table=-11
	/product="B229_C1_175"
	/protein_id="AAI17292.1"
	/db_xref="GI:467110"
	/translation="MTSVIVEIGHTSAEPMLAAAFNGNPGRMPLPTATPHQWLRAVAGGCGHSSARDLAVLRSPAPGRSLASHTSGSFLRQGHSLARGMDGRALVLAGTDEBARADALIGLAADALGRLAAATLRLRGSALAPQLEPAOYDLAVRRWVAELAMAVGCGATVARRAREVELLAOGRVSVRHQVSDVIVLAALCSATERARVVAEALAAATGRLGLPLRWALACLIDIGSVFSEPELSELP"
CDS	1152..1460
	/note="Probable transcription factor (similarity to Streptomyces coelicolor whib); B229_F1_2"
	/codon_start=-1
	/transl_table=-11
	/product="whib"
	/protein_id="AAI17313.1"
	/db_xref="GI:467130"
	/translation="MPQPKOLGPNATIMNMLOGICRGVDSMFPHDGERGRAMOBEORAKEMCRCPVIECCRAHALDGEPTGVGWSEREDLLKGDILARSKISRSA"
CDS	complement(2059..3672)
	/note="GroEL protein; B229_C3_248"
	/codon_start=-1
	/transl_table=-11
	/product="groEL"
	/protein_id="AAI17312.1"
	/db_xref="GI:467130"
	/translation="MSKLIYDETARHAMEVGMNKLADTVRTLGPGRHVLAKAFGPTINDGVTVAREIDLEDPEFNLGALVKSATKTKNDVAGDGTATATVLAQALVKGGLRMVAAGAPVALAGISKAADVSEALLAVATPPVAGKAITQVATVSSRDQICALVGGMNKVGTDVSVSEESTLDTELEFTEGVGFDGFLSAFTVDPSQAVLDDPLVLTQKELISLPLELPLEKVTESGKPLLVADLGEALATLVNSIRTKTLAAVAVKSPFGDRKRAFLDLIAVTGGQVNPETGLVREVGTVLSARVVVSCDPTIIDGSGSNDAAKRVNOLRAIEVDSDEMDREKLOERAVLAGVAVIKGAVETALKRKEVIEDVAAKAKSIEEGLIAGGSGALVCCGALKOIKRTSLTGDEALGIDVFFALKAPLVYATRNAGLDGAVVVDKVSGLPAGHCLANSTIGYGLVDVADVDPYKTVRSVLAASVARMLTTETAVVDKPKATEEHDHGHAF"
CDS	complement(3751..4053)
	/note="10 kd chaperonin; B229_C3_247"
	/codon_start=-1
	/transl_table=-11
	/product="chpn"
	/protein_id="AAI17311.1"
	/db_xref="GI:467129"
	/translation="NAKVKIKPLEDKILVQAGEAETMPSGLVIPENAKERPOEGTVAVGGRMDEGAKRIPVDVSEGDIVYKSGTEIKVNGCEEXLILASRDVAVSK"
CDS	complement(4517..5389)
	/note="Probable glycoproteinase (similarity to Pasteurella hemolytica glycoproteinase A1 and E.coli Y9jC); B229_C3_246"
	/codon_start=-1
	/transl_table=-11
	/product="u229e"
	/protein_id="AAI17310.1"
	/db_xref="GI:467128"
	/translation="MTISVPGCTITLAIETSGDETVGCIACLDVGYTTLADAVASSVDEQARFGGVPEIASSRAHLEALGPTTRCALAAAGLTSKAKDVAAATIGLGAALLVGVAAAKASAAAGVPEYAVNHLGGHLADVYEHGPLECVALLVSGGHTHLLQVRSLGAPIVELGSTVDDAAGEAVDKVARLLGLGYPGCGKVLDDIARGLDGLVAFPGMGTPADLNAFSEFGTKTAAVRYESHDPALPDVAAGPQEAADVADVITMKAVRAATGLVSTLLVGGVAANSRCS"
CDS	complement(5386..6465)
	/EC_number="2.3.1.128"
	/note="30S Ribosomal protein S18 alanine acetyltransferase; B229_C1_170"
	/codon_start=-1
	/transl_table=-11
	/product="rim"
	/protein_id="AAI17291.1"
	/db_xref="GI:467109"
	/translation="MSIVLAIDTAAVTAAGIARFDGHCEFTLAERVTVDAKAHERLTNNVLVALADAEALMCELDAAVVVGGPGPFGRLRYGMATAAYGHALGILPVHXSCLDAIGVRTGDTLVYTPDARRHREYVMARYRQGVRLAGAVSPDIDVDGTLALTVAGSEHAALFLGPLEPILYTPAGLVAAVPMWSVSPILVALYLRRPAKPIITAONEPILVIGTLTPADVRCALQLESQFDDGNPMPAAAFDERLANSYCYGARPADTLVGYATIRLGHTPPFEYEVHTIADVPAVRSGVGRRLGELIDFAGSGAIYLEVRTDNETAIALYRSVGFERIGLRPRYPASGADAVLMRREAQ"
CDS	complement(6462..6947)
	/note="match to 16.9 kd E.coli protein in amib region (y)ey; B229_C2_205"
	/codon_start=-1
	/transl_table=-11
	/product="u229f"
	/protein_id="AAI17300.1"
	/db_xref="GI:467118"
	/translation="MAEYSLRSGAVICERVEDTVAGSLRGEOLRAGDVVLSGFLGAKTVLAKGIAVAMDVDGPIVSPVYLAHRLPRRGTGPAMIHVDDVRLDHRDADLVGLDSDLDITDLAEAVVWVGAGLAECLAAHRLDITRLERVSVDVRIATWQVCSRDLP"
CDS	complement(6940..8001)
	/codon_start=-1
	/transl_table=-11
	/product="B229_C1_169"
	/protein_id="AAI17290.1"
	/db_xref="GI:467108"
	/translation="MKWSPVYAGSSPEPPARPPLTDDIEIRRRNGRPMIAGTGLPVATYVGAISLRISIFERDNACRPDYDRDPEKIGDERRCVITISGMAIYVEEGVAKAPAAVFAIGFYLRMDSPFHQRKRFGRKPGVPRVYDHCGVSSVALDPTLTQLGODLRTVLTQTVPHGMIVLVGSMGILKSPALEVRLTSSRSKLMHSGIASOSLI GPILRAASVSLDRSGIDAFESQRIIMNDTLIAIVLSFLHALEHETAGLWPLRVPA LTAGDHDLTSDERSRGMAVLPILALVIYSGASRLALDQPGAINDGLVRLVRVAPCKALRRRERERLQRHC"
CDS	complement(7931..9097)
	/EC_number="5.1.1.1"
	/note="Alanine racemase; B229_C3_243"
	/codon_start=-1
	/transl_table=-11
	/product="alr"
	/protein_id="AAI17309.1"
	/db_xref="GI:467127"
	/translation="MAVTPISLRPGVLAEAVVDGLAIDYNNRVLEHAGMAQLMVLKADAYGHATQVVALAALAGAELGAVTADALDAIDISAPVLAWLHPGGLDFGPAL IADVOIAYVSROLDLDAVRRGTQRTATVTKATQILNRNGVTDGYDPAMLFTALORA IVEDAVRLRGIMSHLTVADOPNDPNSDQGRKALLAQAHHOGILREFVAHLSNSAT MSRPDLADVVRGTAIVGSLSPVSRGMDGLIPARTVCAVAMVSTRAGEVSYGHD WIAQHDNTNALLPVGADVGFERSIGRLVDYLINGRRRGVGRICMDQGVVDGPGPID VAEQDEALIFLGAGPEPTADQMDLLGTIHVEVTSLRGRITRTYREAOYDVR"
CDS	9105..9446
	/codon_start=-1
	/transl_table=-11
	/product="B229_F3_101"
	/protein_id="AAI17320.1"
	/db_xref="GI:467138"
	/translation="MLFRVCVSTRCHRPRTNGACRSTRVDYGAQLYLMHEVGAVVS AGELIQSPESPQSQTMRSTDTSVASSGCVIGLHLPTECEPFEVFEIELGVPTVSNEQ GGIPGWPQGV"

	CDS	complement(9637..9785) /EC_number="4.1.1.15" /note="Glutamate decarboxylase isozyme; B229_C2_202" /codon_start=1 /transl_table=11 /product="gadB" /protein_id="AAI17299.1" /db_xref="GI:467117" /translation="MGFLAASPDYDCKVAINMGSYPGVGFGVGRNEYGLPEDVER VNYLQAD"
CDS	complement(9787..9951) /EC_number="4.1.1.15" /note="Glutamate decarboxylase isozyme; B229_C1_167" /codon_start=1 /transl_table=11 /product="gadC" /protein_id="AAI17289.1" /db_xref="GI:467107" /translation="MSMPSLSATPTTGELEFPAELCVALLDKLAINDNGDIIVSQAASS GSGCLACTRS"	
CDS	complement(10120..10344) /EC_number="4.1.1.15" /note="Glutamate decarboxylase isozyme; B229_C1_166" /codon_start=1 /transl_table=11 /product="gadB" /protein_id="AAI17288.1" /db_xref="GI:467106" /translation="MAEFDFKTCIDKDDEVPAAAPNSGAAYSMATGLYSRLGTOLRQPV OYLWGVLSPAGMKGDAGLPCHRIALVCSS"	
CDS	complement(10680..11858)	
Query Match	Best Local Similarity 66.4%; Score 1074.7; DB 73; Length 36947; Matches 1278; Conservative 0; Mismatches 318; Indels 3; Gaps 1;	
Oy	1 agcaagctgcgatcgaataacagaacaaccgcggttcggcccataggagtatcgcatgacaag 60	
Db	3669 AGCAAGCTGATTGATACGCACAACCSCCGCTCATGGATGAGTGCGGATGAGAAG 3610	
Oy	61 ctggcgcaaacaccttgtagtgttggggcccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 120	
Db	3609 CTCGCCGACACGCGTCCGGGTAACTTTGGCGCGCGCGGTCCACATGTGTGTGGCCAAG 3550	
Oy	121 gcuitttgccggaaccaagctttaccaaacagcggcgtaacaggttgccaagcttgagattcgaactg 180	
Db	3549 GCATTGGCGGGCCCCACATCACCAAACGAGTGTACACGTCGCGCGCGTAGATGCACCTTA 3490	
Oy	181 gaagatccgattgaagacttggggcgccccagtgtgtgaagtggvtgccccaaaagcaaac 240	
Db	3489 GAGGACCCGTTGAAAACTTGGGCCCTCAGCTGTGTGAAGTGGTGCCACCAAGAACCAAC 3430	
Oy	241 gatgttcgcggttgaagcaccaaccacgccaatcattgycgcgaaggaattcataagggc 300	
Db	3429 GACGTCGCTGGCGAGGGCACTACCAACCCCGACCGCTGTTGGCCCAGGCGCTGTCAGAGGC 3370	
Oy	301 ggccctaagcgttatgtgc 360	
Db	3369 GGCTCGCGCATGATGTGCGCGCGGGGCCNAACCGGTGTGCGCTTAAGTGTGGGATATGACAAG 3310	
Oy	361 gccgcgcgaacggtatatcgaagcgctgctgtgcatalcggccacgcgcggtatccggcaagacc 420	
Db	3309 GCCCGCGACGGCGMTGCCGAGGCACTGCTGGCGGTGAGAACCCCGGTGGCGGGCAAGGAC 3250	
Oy	421 ggcatcgcgaaggtgggcgaagglgtccctcgmgcgagccagatcgyfygactcgtttggc 480	
Db	3249 GCCATCACGCGAGGTGGCAACGGTTTTCTTGCGCGCGAGCAAGATCGGTGGTCTGGTCCG 3190	
Oy	481 gaagcgtgaacaagaagtgcgcacacagacgtygtgtcaaagyttaaagaaatccctccacgtg 540	
Db	3189 GAAGGGATGAACAAGAGTGGCGCACTACGGCGTGTGTCAACGCTTTGAAGAATCGTCCAGCGTGG 3130	
Oy	541 ggcaccaggttggagttcaccgaaggtatlygcttccaaaaagctctctltgcggcalac 600	

Db	3129	GACACCGAGCTAGAGTTTCACCGAGGGCGTCGGCTTCGACAAAGCTTCTCTCGGGCTAT	3070
QY	601	ttcgttcaacgacttcga ttaacacgacgagcggtgtctcagaagaacgctltgatctcgtacac	660
Db	3069	TTTCGAGACCGACTTCGACCTCGCAGCAGAGGCGTGCTCGACGATCCGCTGTCTCGAC	3010
QY	661	caagaacaagatcaagctgtcgtctccgatctgttgcattgtcgtgaagaaggttcagaacg	720
Db	3009	CAGAGAGAGATCACCTCGCTGCTCTGAGTTGCTGCCGATCTCGAAMAGTCACTACGACTCA	2950
QY	721	ggttaagccactacgtatcgtgtgtcttgaaagccgtggagagggcgaaagcglttggaacgttg	780
Db	2949	GGCAAGCGCGCTGATCGTCGTCCGGAAGACTCGAAGGGGAAGCAATTTGGCACCCTTGT	2880
QY	781	gtcaacgcgatttcgaagaagcttgaaagcggtcgcgttcaaaagggcgtaactcgtgtg	840
Db	2889	GTTAACTCGATCCGCAAGACGCTCAAGGGGGTTTGGGTTAACTCCCGCTTCTCGGGGAC	2830
QY	841	cgccgtaagcgatctccttgtaggaacttgcgcgtgtgtgaacggtgtgcacgtgttcaaccc	900
Db	2829	CGAGCGAAGGCGCTTCTCTCGAAGATTTTGGCAATTTGTACCGGGGTGGTCAAGTAGTCAACCCC	2770
QY	901	gaagccggacattgtgtgcgcgcgcgcgtgtggaggtgtcgtgtgtgtgtgtgtgtgtgtgt	960
Db	2769	GAAACCGCGGCTGGTGTGCGCCGAGGGGGGACTGACGTGTGTGGGCTCGGGCCCGGCTGTG	2710
QY	961	gttgttcaagcaagagacgacacgtgtcatgttgcacgcgcgcgcgcgcgcgcgcgcgcgcgc	1020
Db	2709	GTTGTTAGTAAAGACGACGATCATCTGTGACGGCGGGGTTCACAGACGCGGTGTGGCC	2650
QY	1021	aaccggtgcgaaccaacttgtgtgtcgcagatcgcgaacaagaacgattcgatttgatcgtgaa	1080
Db	2649	AAGCGAGTCATCATGCTGTGCGCGCCGAGATTAAGAGGTTCAGGACTTCAGATTTGGATTCGGAA	2590
QY	1081	aagcttggcgaagcgcgctgtgcgaacaactgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	1140
Db	2589	AAGCTGACAGGCGCGCTGTGGCCAACTAGCGCGGTGGGTGGTGGCTGTATTAAGTTGGGGCT	2530
QY	1141	gccaacgcacacgcgcac	1200
Db	2529	GTCACCGAGACGGCGCTGGAAGAAACCAAGAAAGGTCGAAGACGCCGTGGCGGTGGCG	2470
QY	1201	aagcgccggttcgaagagagacgtcgtccctgtgttgaggagagctcgtcattccacacggcc	1260
Db	2469	AAGCGTCTATCGAGAGGGGTATCATCGCGGGGTGGTGGATCGGCGCTCGTCC--AGTGC	2413
QY	1261	cgcaagcgctgcgaacgaacgt	1320
Db	2412	GGCGGGGGGCTCAAGCAATTTGGTCGTACGTGTCGACCGGAGAAAGCCCTGTGCATCGAC	2353
QY	1321	gtgtgttcgcgaagccctgtgcgcgcgtgtgttctgtatctgcgcgcgcgcgcgcgcgcgcgc	1380
Db	2352	GATATTTTTCGAGGGCGCTCAAGGCGACTTGTACTGTGATTGCCACATACAGCCGGGTTGGAT	2293
QY	1381	ggtctcgt	1440
Db	2292	GGCGGGGGTGGTGCAGAGGTTTACCGGGGCTGTGCCCGCGGACAGGACTGAATGCTAGC	2233
QY	1441	aaccttgagctatgtgtgacttggccgctgtgaacggtgtcatctgaacccggttaaggtgactag	1500
Db	2232	ACGGTGGGCTATGGCCGATTTTGGTCGGCGATGTGTGTGTGACCCCGTAAAGTACACGA	2173
QY	1501	tcggcggtgtgtgaacggtgtcatcgtgtgtgcgcggaatgtgttaactcaaacgcgaagagtcgt	1560
Db	2172	TTCACGGTGTCTTAAACGCGGCGCTGGGTGGCGCGGATGTCTTACCACCGAAGACGGCGGTG	2113
QY	1561	gtcgacaagccggtccaaagcagaaga tcaacgacatcaac	1599
Db	2112	GTGGACAAGCGCGGCAAGACAGAAAGACATGTACACACAC	2074

```

U00015/c
LOCUS      42325 bp      DNA                BCT                01-MAR-1994
DEFINITION Mycobacterium leprae cosmid B1620.
ACCESSION  U00015
VERSION    000015.1 GI:466931
KEYWORDS
SOURCE
ORGANISM   .
            Mycobacterium leprae.
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Corynebacterineae; Mycobacteriaceae;
            Mycobacterium.
            1 (bases 38281 to 41190)
            Rinke de Wit,T.F., Bekelie,S., Osland,A., Miko,T.L.,
            Hermans,P.W.M., van Soelingen,D., Drijfhout,J.W., Schoenigh,R.,
            Janson,A.M. and Thole,E.R.
            Mycobacterium contain two GroEL genes. The second Mycobacterium
            leprae GroEL gene is arranged in an operon with GroES
            Mol. Microbiol. 6, 1995-2007 (1992)
            92374850
            2 (bases 2759 to 5248)
            Rinke de Wit,T.F., Clark-Curtiss,J.E., Abebe,F., Kolk,A.H.,
            Janson,A.A., van Agterveld,M. and Thole,J.E.
            A Mycobacterium leprae-specific gene encoding an immunologically
            recognized 45 kDa protein
            Mol. Microbiol. 10 (4), 829-838 (1993)
            95020554
            3 (bases 3688 to 5472)
            Vega-Lopez,F.
            Sequence and immunological characterization of a 42kDa serine-rich
            antigen from M. leprae
            Infect. Immun. 61, 2145-2153 (1993)
            93239328
            4 (bases 1 to 42325)
            Smith,D.R.
            Unpublished
            5 (bases 1 to 42325)
            Robison,K.
            Direct Submission
            Submitted (01-NOV-1993) Department of Genetics, Harvard Medical
            School, 200 Longwood Avenue, Boston MA 02115
            6 (bases 1 to 42325)
            Robison,K.
            Direct Submission
            Submitted (01-MAR-1994) Department of Genetics, Harvard Medical
            School, 200 Longwood Avenue, Boston MA 02115
            On Mar 31, 1994 this sequence version replaced gi:414221.
            This sequence data was produced by the Genome Sequencing Center
            located at Collaborative Research Incorporated (1365 Main St.,
            Waltham MA, 02159). 617-487-7979). Please contact Doug Smith
            (smith@cr.cric.com). The annotation should be considered
            preliminary and incomplete.
            Location/Qualifiers
            1..42325
            /organism="Mycobacterium leprae"
            /db_xref="taxon:1769"
            619..1908
            /note="Trehalose-6-phosphate phosphatase; B1620_F1_1"
            /codon_start=1
            /transl_table=11
            /product="otsp"
            /protein_id="AAC43238.1"
            /db_xref="GI:466950"
            /translation="MPVTIDPRRHSALFDLDAVYVPTPLDSTVTVVROLOGIGVSTA
            VSTNRNSOGVLTATGDLHLFPHVGLASGKTTLVAAANRMAQPGKVVAVDA
            GITAARYGGFALVLGLVIGEDDGHGHDTLRNSGADIVADLGEVIVRTDNRSELDP
            ALQTLGTLADLADPAVPEDEGTLSLDIVDDPSARVPVGALEALOKLATHCPLVIL
            SGRLADVIRKIGVPGIWSGSHGFSTADVGHNDAAEATIPILLEGAAQOLRQOL
            GPVGVAMHEKRGVAVHYRNVARADRVNEVAVAVPAGORNALRVTTGREGVELEPDI
            DMQKGTLMHVIDRLHAGTQGSASIMPLICGDDITDDADPAVAHRTVVGGLPIYVR
            HEDGNRATAALFTLDSPMHVSFETRLARQLSDTOR"
            complement(1979..2188)
            /note="weak match 42.1 kD protein; B1620_C3_270"
            /codon_start=1

CDS
            /transl_table=11
            /product="u1620m"
            /protein_id="AAC43237.1"
            /db_xref="GI:466949"
            /translation="MYCLRGAGFLKPIVLVLSRGVPGRALVLEIGKLAIPNISIY
            LRGVSRDADHDLDLENRARDGVSY"
            complement(2282..2473)
            /note="weak match to (gp:x66077): B1620_C3_269"
            /codon_start=1
            /transl_table=11
            /product="u1620n"
            /protein_id="AAC43236.1"
            /db_xref="GI:466948"
            /translation="MDHRSLSCRDKEGEQEDRCVWPVRYGPDGDAESLJRGPRWS
            SRWGSVGAFTAGGTGFGSGS"
            complement(3910..5136)
            /note="45 kDa serine-rich antigen; B1620_C1_177"
            /codon_start=1
            /transl_table=11
            /product="a945"
            /protein_id="a945"
            /db_xref="GI:466932"
            /translation="MEDFMVYSPENVAFPMRSKPGSTPLMGAAEAMISLAEDLMEAAQ
            EVSDTIVAVPASFAGETSDMLASRVSTFVAMIDGAENAGLIARVLAHVAFAFER
            AGVPLIYLVGNTIHMALKAJNMGCVSTYVAALADYDLMMQNSTAMTYRDTVL
            RENKAMENFEPAPOLYSRCMDRSDVSNFHSSSSDSLTESIDNLTDSVAGSEBGS
            DMSOSYNTGCSVAQSELCDSPEGTGPSQSSQSDSLTSGLTQDLGLDLS1ISSASL
            LTNSSISSSTASSIMPVAVSOVETLGRSOVAVENKIOSSITVAVSDVAQKVVAGV
            GGAVSGALRVPEENMATASQPVWATSHVSPACSAITTVASGLEGTQPAEEVLTA
            VAGSGCTGPAFEEAV"
            complement(5262..5564)
            /note="weak match 32kD protein (sp:pl6645); B1620_C2_218"
            /codon_start=1
            /transl_table=11
            /product="u1620k"
            /protein_id="AAC43225.1"
            /db_xref="GI:466937"
            /translation="MSFPIRVEYGLMAAGRLERTTSFMACNKLTPVTTKVPPA
            ADOVSKLVQSVSSITGKYEGTAQGVDSQSLFVOSLKDPAADQMDSDHMYINTED"
            6340..6816
            /note="transposase; B1620_C3_266"
            /codon_start=1
            /transl_table=11
            /product="u1620a"
            /protein_id="AAC43241.1"
            /db_xref="GI:466953"
            /translation="MPDLCAVNRVYGVKRLNRTAKRAGCHDRHHDIGARVMCLAG
            LQGIYDKRITRTTEHDKQAPRGHVRPDQGLVNAKVPESVESPISRMRFRRCMGY
            TAFCVYFERRIRLPGMGRTSKTTLVLTSLFEDALFTRKYLELRFAMDSAPLGR"
            complement(7286..7606)
            /note="transposase; B1620_C3_266"
            /codon_start=1
            /transl_table=11
            /product="u1620b"
            /protein_id="AAC43235.1"
            /db_xref="GI:466947"
            /translation="MYVHALITTDVNAEGYEETLGGCGDHRRRGRRLVESAADRP
            LAFRRGYLGHQRCRRTARRHNOGQRPVGMQRCRIYLGHPGLRTHGLTVGAVPAT
            LAGV"
            complement(8447..9097)
            /codon_start=1
            /transl_table=11
            /product="B1620_C3_264"
            /protein_id="AAC43234.1"
            /db_xref="GI:466946"
            /translation="MGSRRIRNRLPLMSTPFAWQEFRRDVVVFPGNDFDDCDTV
            DMWGAGGAHPTIDLPVPASSAADDRTDVNDNRGQAEDEKKEPVAFFETNP
            PRVSVAGMDRDIHVELSKVTWMSSESVASLITVLADLRKQASQVYFTDNL
            SQLADDEIRVALILRESVNTNMLSPEDAAEADEVFATRTSDYCPODTENDQM"
            complement(9111..9431)
            /codon_start=1
            /transl_table=11
            /product="B1620_C2_214"

```


[illegible]

RESULT	7					
LOCUS	AR028760					
DEFINITION	AR028760	1620 bp	DNA			29-SEP-1999
ACCESSION	AR028760	Sequence 10	from patent US 5858773.			
VERSION	AR028760.1	GI:5940733				
KEYWORDS						
SOURCE		Unknown.				
ORGANISM		Unknown.				
		Unclassified.				
REFERENCE		1 (bases 1 to 1620)				
AUTHORS		Mazodier, P. and Guglielmi, G.				
TITLE		Regulatory nucleotide sequence of the initiation of transcription				
JOURNAL		Patent: US 5858773-A 10 12 JAN-1999;				
FEATURES		Location/Qualifiers				
	source	1..1620				
		/organism="unknown"				
BASE COUNT		287 a	589 c	540 g	204 t	
ORIGIN						
Query Match		45.9%	Score 741.6;	DB 81;	Length 1620;	

[illegible]

Db 1027 CAGGCGCGCTGCGCCAGATCAAGGCCAGATCGATCGACGATCGGATCGGAGCCG 1086
Qy 1078 gaaaagcttggcagagcggtgcccgaacttggccggggttgcgtgcatacaagtggtg 1137
Db 1087 GAGAAAGCTCCAGAGGCGCCTGCGCAAGCTGCGCGCGGCTGCTGCGATCCGCGTGGC 1146
Qy 1138 gccgcacccgacacgcacacccaaggaagcgaagaaagcgtcagagatgagtgctgcggcc 1197
Db 1147 GCGGCCACCGAGGCTGAGCTGGAAGGAGCGCAGACCGTCTGAGAGAGCGCATTCGCGC 1206
Qy 1198 gccaaagcgcggtcgagagagggatcctccttggtggggaagcctgcataccacag 1257
Db 1207 ACCCGCGCCCGCGGTGAGAGGCGCATGCTCCGCTGGTGGCTCCGCGCTGATCCG 1263
Qy 1258 gcccgaaagcgcttgaccgaactgctgctgcgtgacggcggtgacgggtcctgcgtgc 1317
Db 1264 GCGGTCAAGGCTCTGGAGCAACCTCGGCGC---CGCACCGCGCGAGGCGCACCGGCTGTC 1320
Qy 1318 gacgtgtctcgcgaagcccttgcgcgcgttgtctgtatgcgcgaagcgtgtgtg 1377
Db 1321 GCGGTGCTCCGCGCGCGCGCGCTGAGACCGCTGCGTGTGATCGCGAGAACCGCGGCTC 1380
Qy 1378 gacggtcgtgtgtgtcaacaaggtcagcagctaccgcgcgggcatgggtgtgacgtg 1437
Db 1381 GAGGCTACGTCATCACACCAAGGTGGCGAGCTCGACAAGGCGCAGGCGTTCAACGCG 1440
Qy 1438 aacacctgtgctatgtgtgcttgcgcgtcgagagcggtatcagccgggtcgaagtgatc 1497
Db 1441 GCCACCGCGGAGTACGGCGACCTGTCAAGCGCGGCTGATCGACCGCGTCAAGTCAAC 1500
Qy 1498 aggtcgcggtgtgtgaagcgtatcgtgtgcgcgaatgtactaacaccgaagagctg 1557
Db 1501 GCGTCCCGCTCGAGAACGCGGCGCTCATGCGCTCCCTGCTCTGACGACCGAGACCTTG 1560
Qy 1558 gtgtgtcgaacagcgccgaagcagaagatcaacacatcac 1599
Db 1561 GTGTGCGAAGCGCGCGCGAGAGAGCGCCGAGCGCGTCAAC 1602

RESULT 8
AR028761 2668 bp DNA PAT 29-SEP-1999
LOCUS AR028761
DEFINITION Sequence 11 from patent US 5858773.
ACCESSION AR028761
VERSION AR028761.1 GI:5940734
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 2668)
AUTHORS Mazodier,P. and Guglielmi,G.
TITLE Regulatory nucleotide sequence of the initiation of transcription
JOURNAL Patent: US 5858773-A 11 12-JAN-1999;
FEATURES Location/Qualifiers
source 1..2668
BASE COUNT 446 a 946 c 927 g 349 t
ORIGIN

Query Match 45.0%; Score 727.4; DB 81; Length 2668;
Best Local Similarity 67.3%; Pred. No.1.6e-72;
Matches 1092; Conservative 0; Mismatches 516; Indels 15; Gaps 4;

Qy 4 aagctgatacgaatacgaagaacgcggtcgcgcacatgaggtcgcgcatgagaacgctg 63
Db 855 AAGATCTCGAAGTTGCGAGAGAGCGCCGCGGCTTGAGCGCGCGCTGAGCAACACACTG 914
Qy 64 gccgcaacgctgtcggtgtgagcgtcggtggcgcgcgccgcgcgcatgtgtgtcgccaagcg 123
Db 915 GCGGACACCTTCAGAGCTGAGCCATCGGCGCCCAAGGCGCGCAAGCTGTCATCGAAGAG 974
Qy 124 ttggcgagaccacggtttaccaacgaagcgctacaggtgacagtgagatcagctcggaa 183

Db 975 TTGGGCGCGCCGAGACCATACCAACGAGGCGGTGACCATTCGCCGTGAGGTGCGATGCGAG 1034
Qy 184 gatcgcgttgaagacttggcgcccaagctgtgtgaagtcggttgccacaagaacacgat 243
Db 1035 GACCGCTACGAGAACTTCGCGCGCCACACTGTCAGAGGAGTGGCGGCAAGACCAACGAG 1094
Qy 244 gttgcggttgaagcgaccacacacacacacatcttggcgaagcgatcatcaagggcgcg 303
Db 1095 ATCCGGGTGACGGCACACACACCGGCGATGCTGCGCCAGCGGCTGTCCGCGAGGGC 1154
Qy 304 ctgaggtgaatggcgccgcgcgtcgaacccgctgcgcgtcgcgtgaggttcggaagcgcc 363
Db 1155 CTGGCGAAGCTTCGCGCGCGCGCGCTCCCGCGCGCCCTTGAAGAAGGCGATCGACCGCGC 1214
Qy 364 gccgacggttatcgaagcgctgtgtgcatcgcgcgcgcgtgtgtccgcgaagaccgcgc 423
Db 1215 GTCGCGCGCTGCTCCGCGCGCTGCTCGACACCGCGCGCGCGATCGACGACGATCCGAG 1274
Qy 424 atcgcgcaagtggtgagcgtgtgtcctgcgcgacgagagatgtgtgactggtgtgcgaa 483
Db 1275 ATGCGCGCGCTCGCGCGCTCTCCGCGAGGACAGAGGATGCGGCAAGCTATCGCGGAG 1334
Qy 484 gcatgacaagtgctgcgcgcgaagcgtgtgtgacggtcgaaggaatcctcacgcgtggtgc 543
Db 1335 GCGATGACAAAGTTCGCGGCAAGGAGCGGTGTATCATACGTCAAGGAGTCAACACCTTCGT 1394
Qy 544 accgagtgtagtcaacgaaggtatgtgtctccacaagaagcttctgttcgcatactc 603
Db 1395 GTGCACTGAGCTTACACCGAGGCGATGCGCTTGCACAAAGGCTACCTGTCGCCGTACATG 1454
Qy 604 gttacgcacttgataccacgcagcggtgtctcgagagcggtgtgatcctgcgaacca 663
Db 1455 GTGACCGACCAAGAGCGGTATGAGGCGCTCTCGACACCGCGTCAATCTGTATCCACAG 1514
Qy 664 gacaagatcagctgcgtctccgcatctgttgcacattgttgaaaggttgcgaagcggt 723
Db 1515 GGCAGATTCGCTTCGATCCAGGACCTCTCCGCTGTGAGAAAGTCAATCCAGCGCGGT 1574
Qy 724 -----aagcactactgactgtgtgtcgaagcgttgaagggcgaaagcgttgcgcgcgt 777
Db 1575 GCGTCCAAAGCGCGGTGCTGATCATCGCCGAGGAGGTGAGGCGCGCTGTGAGACCTTG 1634
Qy 778 gtgtcaacgcgattcgcgaagcgttgaagcggttcgagtgcaaggggtccgtactcgt 837
Db 1635 GTGCTCAACAAAGATCCGCGCGCATGTCACGCGCTCCCTCAAGGCGCGCGCTTCGT 1694
Qy 838 gaccgcgtaagcggttcccttgaggaacttgcggtgtgtgtgcggttgcgcgtgtgtcaac 897
Db 1695 GACCGCGCAAGGCGATGCTCGGCGACATGCGCACCTCACCGGTGCGCATCTATCGCC 1754
Qy 898 cccgaacgcgcacttgcgtgcgcgaagtggtgttgagtgctgtgagtcgtgcgcgcgaagc 957
Db 1755 GAGGAGGTGCGGCTCAAGCTCGACGACGCGCGGTGAGTGTGAGGCAACGCGCGCGCG 1814
Qy 958 gtgtgttgaagcaaggaagcagcgttcatgttgcagcggtgcgcgcgcgcgaagcggt 1017
Db 1815 GTCAACGCTCACCAAGGAGCGACGACCATGTCGACTTGAGAAAGACGCGCGAGGAGTTC 1874
Qy 1018 gccaaacggcggaacacactgctgtccgagatctgaacaaggagattcggatttggatcgg 1077
Db 1875 CAGGCGCGCGTCCGCCAGATCAAGCGCGGAGATGAGTGCAGCCGACTGGGATGGAGCGC 1934
Qy 1078 gaaaagcttggcgaagcgtgtgcgaacttgcgcgcgggtgtgtctlcaacaaggttggt 1137
Db 1935 GAGAAGCTTCAGAGAGCGCGCTCGCCAAAGCTTGGCGGCGCTGTGCGTATCCGCTCGCG 1994
Qy 1138 gccgcacacgaacacgcactccaagggagcgcaaggaagcgctgaagatgtgtgtcggcgc 1197
Db 1995 GCGGCGCACCGAGGTTCAGCTGGAAGGAGCGCAACACCGTCTGGAGAGACGCGCATCTCGCG 2054
Qy 1198 gccaaagcggttcgagagggcatcgttccctgtgtgtgggaagcctgcgtcatcaccag 1257

```

Db 2055 ACCCGCGCCGCGTGCAGAGAGGAGCATGCTCCGGTGAGCTCCGCGCTGCTCAC--- 2111
Oy 1258 gcccgcaagcgctgaccgaactgctgctgctgctgacggcgtaagctcctgctgctc 1317
    ||| ||||| ||| ||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2112 GCCGTCMAAGTCTGTGACGACACACTCG---GCCGACCGCGGAGAGAGGCCCGGTGTC 2168
Oy 1318 gacgctgtcttcgcaagcccttcgcgcgcgctgtctgtatgcgcgcgaactgctgtc 1377
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2169 GCGGTCGTCGCCGCGCGCGCTCGAGCGCGCTGATCGCGAGAACCGCGGCTC 2228
Oy 1378 gacgctgctgtgtgtgtaacaagctcagcgactaccgcccgcgcatgaggtgtaactg 1437
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2229 GAGGGCTACGTATACACCAAGGTGGCGGAGCTCGAAGAGGCGAGGCTTCACGCG 2288
Oy 1438 aacaccctgagctatgctgacttgccgctgacgcgctgcatcgacgcgtcaagtgact 1497
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2289 GCCACCGCGGAGTACGCGGCACTGGTCAAGCGCGCGGCTGATGACCGGTCAAGGTACAC 2348
Oy 1498 aggtcgcggtgtgtgacgcgctgcatcggttgcgcgagtgtactacacacagagacggtc 1557
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2349 CGCTCCGCCCTGAGAGACGCGGCGCTCAATCCGCTCCGCTGCTGACGACGAGACCGTG 2408
Oy 1558 gtgtgcgacgaagcgcgcaagcgagaagatcagacatcac---caaggcgacgcgcac 1614
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2409 GTGTCGAGAAACGCGCGAGAGAGAGCCGCGGCTGACGCTACGCGGACAGCCAC 2468
Oy 1615 tga 1617
    |||
Db 2469 TGA 2471

```

RESULT 9

```

LOCUS SCG64 41055 bp DNA BCT 28-MAR-2000
DEFINITION Streptomyces coelicolor cosmid 664.
ACCESSION AL031317
VERSION AL031317.1 GI:3449234
KEYWORDS 10 kb chaperonin; 30S ribosomal protein S11; 30S ribosomal protein
S13; 30S ribosomal protein S9; 50S ribosomal protein L13; 50S
ribosomal protein L17; 50S ribosomal protein L36; 60 kD chaperonin;
acetyltransferase; acps; adenylate kinase; adk; alanine racemase;
alt; coaA; cpn10; cpn60; dehydrogenase; DNA-directed RNA polymerase
alpha chain; gcp; gmsI; glucosamine-1-fructose-6-phosphate
aminotransferase; groEL; groES; holo-lacyl-carrier protein]
synthase; infA; lipase; map; methionine aminopeptidase;
O-sialoglycoprotein endopeptidase; pantothenate kinase;
phospho-sugar mutase; pseudouridylylate synthase; rplM; rplO; rplJ;
rpoA; rpsI; rpsK; rpsM; translational initiation factor IF1; trnA;
two-component regulator.

```

SOURCE

```

ORGANISM Streptomyces coelicolor A3(2).
Bacteria; Firmicutes; Actinobacteria; Streptomycetaceae; Streptomyces.
Actinomycetales; Streptomycinae; Actinobacteridae;

```

REFERENCE

```

AUTHORS 1 (bases 1 to 41055)
Redenbach,M., Kieser,H.M., Denapalte,D., Eichner,A., Cullum,J.,
Kinasli,H. and Hopwood,D.A.
TITLE A set of ordered cosmids and a detailed genetic and physical map
for the 8 Mb Streptomyces coelicolor A3(2) chromosome
Mol. Microbiol. 21 (1), 77-96 (1996)

```

JOURNAL

```

MEDLINE 97000351
REFERENCE 2 (bases 1 to 41055)
AUTHORS Saunders,D. and Harris,D.
JOURNAL Unpublished
AUTHORS 3 (bases 1 to 41055)
TITLE Direct Submission
PARKHILL,J., BARRELL,B.G. and RAJANDREAM,M.A.

```

```

COMMENT Submitted (19-AUG-1998) Streptomyces coelicolor sequencing project,
Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.
David A. Hopwood, [3] John Innes Centre, Norwich Research Park,
Colney, Norwich, Norfolk NR4 7UH, UK
Notes: Streptomyces coelicolor sequencing at The Sanger Centre is funded

```

FEATURES

source

by the BBSRC. Details of S. coelicolor sequencing at the Sanger Centre are available on the World Wide Web. (URL: http://www.sanger.ac.uk/Projects/S_coelicolor/) CDS are numbered using the following system eg SC7B7.01c. SC (S. coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary strand).

The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous. The length in codons is given for each CDS. Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database. The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon. Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(22):4768-4778(1994)) and the FramePlot program of Bibb et al., Gene 30:157-66(1984) as implemented at <http://www.nih.gov/jp/jun/cgi-bin/frameplot.pl>. CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, tgg or (att)) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.

IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions. Cosmid 664 lies between D31 and D63 in the AseI-D genomic restriction fragment.

Location/Qualifiers

```
1. .41055
```

```
/organism="Streptomyces coelicolor A3(2)"
```

```
/strain="A3(2)"
```

```
/db_xref="taxon:100226"
```

```
/clone="cosmid 664"
```

```
1. .1610
```

```
/note="98.4% match to SCSECYDNA X83011 S.coelicolor secY
```

```
to cus DNA from 4547 to 6154"
```

```
1. .286
```

```
/gene="adk"
```

```
<1. .286
```

```
/gene="adk"
```

```
/note="SCG64.01, adk, adenylate kinase, partial CDS, len:
```

```
> 94 aa; almost identical to KAD_STRGO adenylate kinase
```

```
(EC 2.7.4.3) (205 aa) and highly similar to many e.g.
```

```
KAD_BORPE adenylate kinase (EC 2.7.4.3)(218 aa), fasta
```

```
scores: opt: 2.53 z-score: 429.1 E(): 1.2e-16, 50.6%
```

```
identity in 85 aa overlap. Contains Pfam match to entry
```

```
PF00406 adenylatekinase, Adenylate kinases, score 136.80,
```

```
E-value 3.2e-37"
```

```
/codon_start=2
```

```
/transl_table=11
```

```
/label="adk"
```

```
/product="adenylate kinase"
```

```
/protein_id="CA20379.1"
```

```
/db_xref="GI:3449235"
```

```
/db_xref="SPTREMBL:086771"
```

```
/translation="IAGRRVCRNEPKHVFHYTPPKKEGVCDCGGLYQRDDSDSE
```

```
TVRRLEVYHTQREPIIDYKSGLVATIAATGPDVTPRALEALKRQ"
```

```
2. .208
```

```
/gene="adk"
```

```
/note="Pfam match to entry PF00406 adenylatekinase,
```

```
Adenylate kinases, score 136.80, E-value 3.2e-37"
```

```
/label="adk"
```

```
/note="possible RBS upstream of map"
```

```
433. .426
```

```
/label="adk"
```

```
/label="adk"
```

```
/label="adk"
```

```
/label="adk"
```

```

/note="SC66d_02, map, methionine aminopeptidase, len: 278  

aa; identical to TR:O54208 (EMBL:X83011) methionine  

aminopeptidase (278 aa) and highly similar to many  

e.g. MPM_BACSU m ethionine aminopeptidase (EC 3.4.11.18)  

(248 aa), fasta sco res: opt: 546 z-score: 922.2 (E): 0,  

44.5% identity in 256 aa overlap. Also similar to S.  

coelicolor map2 (E): 3.2e-3, 46.8% identity in 267 aa  

overlap). Contains P500680 Meth ionine aminopeptidase  

subfamily 1 signature and Pfam match to entry PF00557  

pep_M24, metalloproteinase family M24, scor e 194.80,  

E-value 1.4e-54"  

/codon_start=1  

/transl_table=11  

/product="methionine aminopeptidase"  

/protein_id="CAA20380.1"  

/db_xref="GI:3449236"  

/db_xref="SPTREMBL:O54208"  

/translation="WQIKNEPQIAKKMEAGLVVAATHATREAAPVATTKLDYVA  

RKVLVAENHAKPNFLGYGGEPAITCTSVNEVVNHGIPSDVVLKGDVLSIDCGAIIQGG  

MHGDAVAFVSGSHSPELVELSESLRVTESSMAGALNGLVDVRALERTYLRDP  

RKGCKAYGIEDYGSGHGIEKHMDPHLLNVDRRGKGPRLVGFCAIPLPWISLTGS  

PREVLPRPEMWYTITDGTWSSHHEHVALLDGCPVLVLTSPGGRAKLAELGITNAPRR  

LA"  

misc_feature  

444..1196  

/gene="map"  

/note="Pfam match to entry PF00557 pep_M24,  

metalloproteinase e family M24, score 194.80, E-value  

1.4e-54"  

'label=*  

'gene="map"  

'note="PS00680 Methionine aminopeptidase subfamily 1  

signature"  

'/label=*  

1405..1408  

/note="possible RBS upstream of infA"  

'/label=*  

1416..1637  

/gene="infA"  

1416..1637  

/gene="infA"  

/note="SC6c4_03, infA, translational initiation factor  

IF1, len: 73 aa; identical to TR:O54209 (EMBL:X83011)  

translat ional initiation factor IF1 (fragment) (65 aa)  

and highly s imilar to many e.g. IF1_BACSU translation  

initiation factor IF-1 (71 aa), fasta scores: opt: 384  

z-score: 792.5 (E): 0 , 78.6% identity in 70 aa overlap.  

Contains Pfam match to e ntry PF00575 SI, S1 RNA binding  

motif, score 54.10, E-value 3.1e-12"  

/codon_start=1  

/transl_table=11  

'/label=infA  

'/product="translational initiation factor IF1"  

'/protein_id="CAA20381.1"  

'/db_xref="GI:3449237"  

'/db_xref="SWISS-PROT:O54209"  

'/translation="MAKKOGALEITEGTVVESLPPANMEKVLELNQHVLANSKMRNH  

YRIILPRDVAVELSPLYLTRGRIVYRK"  

1422..1634  

/gene="infA"  

/note="Pfam match to entry PF00575 SI, S1 RNA binding  

motif , score 54.10, E-value 3.1e-12"  

'/label=*  

1690..1693  

/note="possible RBS upstream of rpmJ"  

'/label=*  

1706..1816  

/gene="rpmJ"  

/note="Pfam match to entry PF00444 L36, Ribosomal protein  

L36, score 70.40, E-value 3.1e-22"  

'/label=*  

1706..1819

```

[illegible]

Db 34630 GTCGTCAACAGATCCGGGACCTTCAACGGGCGTGCATCAAGCCCGCCGCTTCGGC 34689
Oy 838 gaccgcgtaagcgcttcccttgaggaccctgagcggtgagcgagtgagcgagtgagcag 897
Db 34690 GACCCCGCAGAGGCGATGCTGAGAGACATGCGCTTCACCGGGCCACGGTCAATTC 34749
Oy 898 cccgagcgccgactgtgtcgcgaggttgagcggtgagcggtgagcggtgagcgagcg 957
Db 34750 GAGGAGGTGCGGCTCAAGCTGACAGAGGTGCGGCTGAGGTCCTGCGACCGCCGCGC 34809
Oy 958 gtgtgtgtcgaagaagagacagcgatcattgtcagcgcgcgagcgagcgagcgagcg 1017
Db 34810 ATCACCCTGACACAGAGAGACACCAAGATCGTCGAGCGGCGGCAACGCGCACGAGT 34869
Oy 1018 gccaacgcgagcgagcgagcgactgtgtcgcgagatcgacgaagaagcgatcgattggatcg 1077
Db 34870 CAGGCGCCGATCGCCCAATCAAGCCCGAGATCGAGAACAGGACTCCGACTGGGACCGC 34929
Oy 1078 gaaaagcttgcgagcgagcggtgagcgagcgagcgagcgagcgagcgagcgagcgag 1137
Db 34930 GAGAGGCTTCAGAGAGCGGCTCGGAGAGTGGCGGGCGGCTGCTGATCAAGGTCCGC 34989
Oy 1138 gccgcgacgagcgagcgagcgactcgaagaagcgagcgagcgagcgagcgagcgagcg 1197
Db 34990 GCCCGCACGAGAGTGAAGGAGCGCAAGCAGCAGCGTGTGAGAGCGCCATCTCCGC 35049
Oy 1198 gccaacgagcgagcgagcgagcgactcgtccctgtgtgagcgagcgagcgagcgagcg 1257
Db 35050 ACCCGCGCGCGCTGAGAGAGGCGATCTCTCTCGGTGCTGCTGCTGCTGCTGCTGCT 35109
Oy 1258 gccgcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcg 1317
Db 35110 GTCAAGGTGCTGAGAGGCAACTCGG-----CAAGACCGGCGAGAGCGCCACCGGTGC 35163
Oy 1318 gacggtgtctcgaagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcg 1377
Db 35164 GCGGCTGCTCGCGCGCGCGGCTGAGAGGCGCGCTGAGATGCGGAGAACCGCCCGCTG 35223
Oy 1378 gacggt 1437
Db 35224 GAGGTTACGTCATCACTCCAGGTCGCGGACCTCGACACTCGACAGAGGCGCGCTTCACGCGC 35283
Oy 1438 aacacgctgaagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 1497
Db 35284 GCCACCGCGAGTACGAGGAGCGACTGCTCAAGGCGCGGCTGATGACCGCGTGAAGTACAC 35343
Oy 1498 aggtcgt 1557
Db 35344 CGCTCGCGCTGTGAGAACGCGCGCTTCATCGCTCTCTGCTGACGACCGACCGCTG 35403
Oy 1558 gtgtgtcgaagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcg 1615
Db 35404 GTCTGTGAG 35461

RESULT 10
MSGBCG 2431 bp DNA BCT 26-APR-1993
LOCUS M.bovis BCG gene encoding antigen A (MbaA), complete cds.
DEFINITION M17705
VERSION M17705.1 GI:149933
KEYWORDS antigen.
SOURCE M.bovis (strain BCG P3) DNA, clones PRIB110[2,3].
ORGANISM Mycobacterium bovis
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacteriaceae; Mycobacteriaceae;
Mycobacterium; Mycobacterium tuberculosis complex.
1 (bases 1 to 2431)
Thole,J.E.R., Keulen,W.J., Kolk,A.H.J., Groothuis,D.G.,
Berwald,L.G., Tiesjema,R.H. and van Embden,J.D.A.
Characterization, sequence determination, and immunogenicity of a
64-kilodalton protein of Mycobacterium bovis BCG expressed in
Escherichia coli K-12

JOURNAL Infect. Immun. 55, 1466-1475 (1987)
MEDLINE 87193155
FEATURES
Source Location/Qualifiers
1..2431
/organism="Mycobacterium bovis"
/db_xref="taxon:1765"
576..2198
/note="antigen A"
/codon_start=1
/transl_table=1
/protein_id="AA025358.1"
/db_xref="GI:149934"
/translation="MAKTIAYDEBARRGLEKLNALADAVKVTLGPKGRNVLEKKWG
APRTINDGVSIAKEIELEDDPEYKIGALIVEKAKTDDVAGDGTATVLAOLREG
LRNYAAGANPLGLKRGIEKAEKVEKTEFLILGAEKVEKEDVATATASAGDSIGDI
AEAMDKVSNECVITVESNFTGDLLETECHMRPKKTSISYFTVDPERQAVLEDPYI
LVSSKSTVKDLPLEKVIYGAERPLITIAEDVEGASLTLVNKLRTGFFSVAVNA
PGFDRRKAMLDMAILTTGQVISEEVLTLENADSLGKARKVVTKDETTVEGA
GDNDIAIRVAOIRIOEINSDYDREKLDERLAKLGAIVAKAGAAVELEKHK
RIEDAVRNAAAVEGIVAGGVTLLQAPLTDLEKLEGEATCANIVKVALEAPLKO
IAFMSGLEPGVIAEKVRNLPAGHGLNMGTVYEDLLAAGVADPVKVTIRSLAQNAASIA
GIPLTFAVVDAREKREKASVPGSGDNGHDF"
BASE COUNT 446 a 736 c 878 g 371 t
ORIGIN 2 bp upstream of MbaA site.
Query Match 43.9% Score 709.8; DB 2; Length 2431;
Best local similarity 66.0% Pred. No. 1.5e-70;
Matches 1048; Conservative 0; Mismatches 532; Indels 9; Gaps 1;
Oy 3 caagcttcgaatacgaagaacgcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcg 62
Db 581 CAAGCAATTTGGGTACGACGAAGAGGCGCGGCGGCTCGAGCGGGCTTGAACGCCCT 640
Oy 63 ggcgcgacacggtgtgcggtgtgcggtgtgcggtgtgcggtgtgcggtgtgtgtgtgtgtgtgt 122
Db 641 CGCGCATCGGTAAAGGTGACATTGGGCGCCCAAGGCGCGCAACGTCGTCTGGAAGAA 700
Oy 123 gtttggcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcg 182
Db 701 GTGGGCTCCCCACGATACCAACGATGTGTGTCTCATCGCCCAAGAGGATCGAGCTGGA 760
Oy 183 agatcgttgaagaacttgcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 242
Db 761 GGATCGGTACGAGAAAGATCGGCGCGGAGCTGTGTAAGAAGGTAGCCAGAAAGACCGATGA 820
Oy 243 tgttgcggtgaagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 302
Db 821 CGTCGCGGTGACGACCAACGAGCGGCAACCGTGTGCGCCAGCGGCTTCCGAGAGCG 880
Oy 303 ccttgaagcttaagtcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 362
Db 881 CTTGCGCAACGTGCGGCGCGGCGGCGCAACCGCTGCTTCAACGCGGCAATCGAAAGAGC 940
Oy 363 cgcgcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcg 422
Db 941 CGTGAGAGAGGTACACCGAGACCTGTCTCAAGGCGCGCCAGAGAGGTGACAGACCAAGAGCA 1000
Oy 423 catcgcgaggtgtgcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 482
Db 1001 GATTGCGCGCACCGCACGAGATTTGCGCGGTGACAGACGATGCTGACCTGATCGCGCA 1060
Oy 483 agcgatgaagaagtcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 542
Db 1061 GCGATGAGCAAGGTGGGCAACGAGGCGGTATCATCACTGACAGAGATCAACACCTTTGG 1120
Oy 543 caccgagttgagttcaccgaaggtatgtgtcctcacaagaagccttctgtgcgagcattt 602
Db 1121 GCTGAGCTGACGCTACCGAGGAGGTATGCGTTGACAGAGGCTACATCTCGGGGTACTT 1180
Oy 603 cgttacgagctcataacagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 662
Db 1181 CGTGACGAGCCCGAGAGGCTCGAGAGCGGCTCTGAGAGACCCCTACATCTGCTGTCAG 1240

[illegible]

Db	677	GATTTCGGCCACCCGACGATTTTCGGCGGGTGAACCACTCATCGTGAACCTGATCCGCCA	736
QY	483	agcagatgacaaggctcgcgcacgacgacgctgctcgaagcgtctcaagaaatccctcaacgctlgg	542
Db	727	GGCGATGACAAGGTGGGCAAGGGGCGTCATCACCGTCGABGAGTCCAAACACTTTGG	796
QY	543	caaccgagttgagttacgtccgaaggtatctgcttccaaaggctctctgctggcaactt	602
Db	797	GCTGAGCTCAGCTCACCAGAGGATTAGCGGTTGCAAGAAGGCGTACATCTCGGGCTACTT	856
QY	603	cgttacgacacttgataataacgaagcggtgctccgagcgcgcttgatcgtctatcacca	662
Db	857	CGTGACCGACCCGGAGCGTACGAGGGGGGCTCTGGAGGAGCCCTTCATCTCTCTGTGTAG	916
QY	663	agacagaatlcagctcgtctcccgatctgttgcattgtctggaagaaqgttgcagaaacgg	722
Db	917	CTCCAAAGTGTCCATCTGTCMAAGGATCTGCTGCCGCTGCTCAGAMAGGTATCTGGAGCGG	976
QY	723	taagcgaactatgactcgttgcgttgaagaacgttgaagggcgaagcgtttgcgcgcgttgcgt	782
Db	977	TAAAGCCCTGCTGATCTATCCCGCGAGCGCTGAGGGCGAGCGGCTGATCCACCTCTGGTGT	1036
QY	783	caacgcgatttcgacaagcgtttgaagcggtgcggtgtcaagggccgctacttcgtgtgacg	842
Db	1037	CACCAAGTATCCGGCGCACCTTCAAGTGGTGGGGGTCAAGAGCTCCCGGCTGGGAGCACG	1096
QY	843	ccgttaagcgtcttcccttgaagccttgcggtgcgtgtgaagggctgacacgcttgttcaaccgga	902
Db	1097	CCGCAAGGCGGATGCTGAGAGGATATGGCGCATTTCCACCGGTGGTCAAGTATACGGGAAGA	1156
QY	903	cgcgcgcaattgtctctgcgcgaagltgagcgttgaaggtgtcttggactcggccgacgcgtgtgt	962
Db	1157	GGTGGCGCTTGACGCTGGAGAACCCCGCACCTGTGCTGTAGGCAAGGCGCCGCAAGTGTGT	1216
QY	963	gttcaagcaaggaagaacacggttcatatttgcagcgcgcgcgcacacgcgaagacggttgcgaa	10222
Db	1217	GGTCAACAAGAGACGACACCACTGTCGAGGGCGCGGTGACACCGAGCGGCATGTGCCGG	1276
QY	1023	ccgcgcgaacaactctgcgttgcgcagatlcgaacaagaacgcatctcgatcttggatctgcygaaa	10828
Db	1277	ACGAGTGGCCCATGATTCGCCAGGAGATTCGAGAACACGCACTCCGACTACGACCTGTGACAA	1336
QY	1083	gcttgcgcagcgcgtcttgcgcaaaacttgcgcgcgcgcgcgcgttgccttcatcaagttgtgcgc	11422
Db	1337	GCTGACGAGAGCGCTGGCCAAAGCTGGCGCGGTGCTGCGGGTGATCAAGCGCCGCTGCCG	1396
QY	1143	caaccgaacccgcatctaaagagcgcgaagaagcgttcgaaggttgcgttgcgcgcgcgcgaa	12022
Db	1397	CACCGAAGTCAACTTCAAAGGAGGCGCAAGCACCCGATCGAGATGGCGTTTCGCATATGCCAA	1456
QY	1203	ggccgcgcgttcgaagaaggaatcttcccttgcgttggggagacctcgtctcatccacaagccgc	12622
Db	1457	GGCCGCGGTGCAAGGAGGCGATTCGTCCGCGGTGGGGTGTGACGCTGTTTGCAAGCGGCCCC	1516
QY	1263	caagcgctgcacgcaactcgtgtgcgtgcgtgcacgcgttgaagaggttccctgcgttgcgcagct	13222
Db	1517	-----GACCTTGACACAGCTGGAAGCTCGAAGGCGAGAGGAGACCGCGCCACCAT	1567
QY	1323	gttctccgaagcccttgcgcgcgcgcgttgttcttgatctgcgcgcgcgaacgcttgccttggacg	1382
Db	1568	CGTGAAGTGGCGCTGGAGGCGCCCGCTGGAAGCAGATGCGCTTCAACTCCGGGCTGGACCC	16277
QY	1383	ctcgcgtgtgtgttcaacaagaatlcagcgcgattcaaccgcgcgcgacatctggaacgttgaacac	14422
Db	1628	GGGCGGTGGTGGCCAGAAGTGTGGCAACTCGCCGCTGGGCCACGAGCTGAAGCTCAACAC	16877
QY	1443	ccctgagctatgttgaacttgcgcgttgcgcgttcatcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	15022
Db	1688	CGGTGCTACAGAGATGTGCTGCTGCGCGCGCTGTGCTGACCCGGTCAAGGATGACCCGTTTC	17477
QY	1503	ggcgcgttlttgaacgcgttcatcgcgttgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	1552
Db	1748	GGCCCTCGAAGATCCGGCGTCAATCGCGGGGGCTGTTCTCTGACACCGAGGCGCGTGTGTC	18077

Query Match	43.9%	Score 709.8	DB 81	Length 4380
Best Local Similarity	66.0%	Pred. No. 1.1e-70		
Matches 1048; Conservative	0	Mismatches 532	Indels 9	Gaps 1

[illegible]

KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
REMARK
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Mycobacterium tuberculosis.
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacterineae; Mycobacteriaceae;
Mycobacterium; Mycobacterium tuberculosis complex.
1 (bases 1 to 19781)
Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C.,
Harris, D., Gordon, S.V., Eigmeier, K., Gas, S., Barry III, C.E.,
Tekala, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T.,
Connor, R., Davies, R., Devlin, K., Fellwell, T., Gentles, S.,
Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J.,
Moule, S., Murphy, L., Oliver, S., Osborne, J., Quail, M.A.,
Rajandream, M.A., Rogers, J., Rutter, S., Seeger, K., Skelton, S.,
Squires, S., Squires, R., Sulston, J.E., Taylor, K., Whitehead, S., and
Barrell, B.G.
Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence
Nature 393 (6685), 537-544 (1998)
98295987
Erratum: [[published erratum appears in Nature 1998 Nov
12;396(6707):190]]
2 (bases 1 to 19781)
Parkhill, J.
Direct Submission
Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium
tuberculosis sequencing and mapping teams: Sanger Centre, Wellcome
Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique
Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux,
75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk
On Jun 27, 1998 this sequence version replaced gi:2909511.
Notes:
Details of M. tuberculosis sequencing at the Sanger Centre are
available on the World Wide Web.
(URL: <http://www.sanger.ac.uk/Projects/M.tuberculosis/>) CDS have
been renumbered from the original cosmid submissions but the old
gene designations are in brackets after the new gene numbers.
Gene prediction was based on a Hidden Markov Model of TB genes
implemented in TBPbase (Krogh) supplemented with visual inspection
of positional base preference in codons, especially where there is
an increase in the observed/expected third position G + C.
CAUTION: In some cases we may not have predicted the correct
initiation codon. Where possible we choose an initiation codon
(atg, ggg, or ttg) which is preceded by an upstream ribosome
binding site sequence (optimally 5-13bp before the initiation
codon). If this cannot be identified we choose the most upstream
initiation codon.
location/Qualifiers
1. 19781
/organism="Mycobacterium tuberculosis"
/strain="H37RV"
/db_xref="taxon:1773"
<1. >19781
/note="fragment designated v037. Does not represent a
physical clone"
complement(42..1259)
/gene="moaA3"
complement(42..1259)
/gene="moaA3"
/note="RV0438C, (MTV037.02C), len: 405. moeA-2, probable
molybdenum cofactor biosynthesis protein (moeA-2) similar
to many e.g. AN10817_2 (429 aa) and to Mycobacterium
tuberculosis MoaA; MTC1237.08 (426 aa). FASTA scores:
Y10817|AN10817_2 A.nicotlinovora moaA, moeA, moeC (429
aa) opt: 786z-score: 843.3 E(): 0; 39.2% identity in
398 aa overlap; and Z94752|MTC1237_8 (426 aa) opt: 667
z-score: 716.7 E(): 2e-32; 36.5% identity in 425 aa
overlap. TBPbase score: 10.889"
/codon_start=1
/transl_table=11
/product="moaA3"
/protein_id="CA117395.1"
/db_xref="GI:2909513"

/db_xref="SPRREMBL:053725"
/translation="MRSVOEHQHVVAEMNRAQRPITVPLTQAGLIVGGDVVAPLSIP
VFNNSAMDCYAVAEPTSSGATPQNPVPLVEADIEIPGRADMILTQPVTRHRTMTGAPV
PTEAIVPEEATDGGVDSVAIRQQAETPKKHTRSEEDVAGCTVLAHQVITPVVLG
LAAALILAEPLVLRQRLVITSGSELSAIPGPILOGQIYESNSIMLAADVADGAAV
VATATGDDVAGAILDRVADADILITSGSVASAGAEVYKAFESADYRGDGHVE
FVKVAMPQPGQGVAGVACPITVTLTGNFVSALVSEFETRLPRLRMALPLPPEPRH
SAVLTASLTPSPKRRQFRRAILDHQAGTVISGPPASHHLRMILASANGLLDIPEDVE
VAGPQLOVWDLT"
complement(1278..2213)
/gene="RV0439C"
complement(1278..2213)
/gene="RV0439C"
/note="RV0439C, (MTV037.03C), len: 311. Probable
oxidoreductase similar to several e.g.
SPI003326|OXIR_STRAT (298 aa). FASTA scores:
A40116|A40116_1 Sequence 6 from Patent WO9423043. U (320
aa) opt: 518 z-score: 605.3 E(): 3.2e-26; 35.5% identity
in 304 aa overlap; and SPI003326|OXIR_STRATPROBABE
OXIDOREDUCTASE (EC 1.-.- (298 aa) opt: 400 z-score: 469.6
E(): 1.2e-18; 34.6% identity in 298 aa overlap. TBPbase
score is 0.903"
/codon_start=1
/transl_table=11
/product="hypothetical protein RV0439C"
/protein_id="CA117396.1"
/db_xref="GI:2909514"
/db_xref="SPRREMBL:053726"
/translation="MTARNDKTRKMSADAVPDQSGRVVVVTCGANTGIGYHTAVFADR
GAHVAVLAVRLKNGAARIMARPGAVTLOQLDCLSDSVRAADALRTAYPID
VLINAGVMMTPKQVTKDGELOFGTNHGHFALTLGLVDHMLPVGSAVVVYVSSGSH
RIHAATHPDIDQWERXRYNAYCAKILNLFYVLOQRICGACKSTTAVAAHPCGS
MELTIRNLRLIPVATVIGPLIPSPKMGALPTLTAADPTTGGQVITGPDSFGQR
GHPKVVQSSAQSHDKDQRRMTVSELTGVSFGV"
2490..2495
/note="possible RBS, 99agga, for RV0440"
2507..4129
/gene="groEL2"
2507..4129
/gene="groEL2"
2507..4129
/note="RV0440, (MTV037.04), len: 539. groEL2, identical
to M. tuberculosis GroEL-2. Has P500296 Chaparonsins cpn60
signature. FASTA scores: SPI06806|CH62_MCTU 60 KD
CHAPERONIN2|PROTEIN CPN60 (539 aa) opt: 3318 z-score:
3673.1 E(): 0; 100.0% identity in 539 aa overlap. TBPbase
score is 0.847"
/codon_start=1
/transl_table=11
/product="groEL2"
/protein_id="CA117397.1"
/db_xref="GI:2909515"
/translation="MAKTAYDEEARGLERGLNALADAVKVTLLGFGKRVNVELEKMG
APTITNDGVSIAKEIELEDPYEKIGAEIVKAEKVDVAGDGTTPAVLQALVREG
LRNVAGVADPLIKRGIEKRAVEKVTETLLKAGKVDVAGDGTTPAVLQALVREG
IARMDKVGNDGVTVTEESNFGQLLETGCMRPDKCYISGVPVTPDEQEAVIDEYI
LVYSKVSSTVTKDLPLEKVIYGGKPLTIYAEVDEEALSTLVYKTRKTPSVAYKA
PGFGDRKAMLODMALITGGVYISEEVGLTLEAVDLSLIGKAKVVTDEDTTVEGA
GDIDATAGVAAQTRQEIENSDDSDYDEKLOEIRLAKAGVAVIKAGAAVELEKRR
RIEDVARNKAAVEEGIVAGGVITLQAAPTLDELKLEDEATGAVIVALEAPLKO
IAFNSGLPEGVAAEKYRNLPAHGILNAGMGYVEDLLAAGVADPVKVTSRALQNAASIA
GFLTTTEAYVADKPEKRSVROGGMGMDP"
3713..3748
/gene="groEL2"
/note="P500296 Chaparonsins cpn60 signature"
complement(4195..4623)
/gene="RV0441C"
complement(4195..4623)
/gene="RV0441C"
/note="RV0441C, (MTV037.05C), len: 142. Unknown but
identical to hypothetical M. tuberculosis protein
MSRCWPA_4 (142aa). FASTA scores: M15467|MSGTCWPA_4
M.tuberculosis 65 kDa antigen (ce (142 aa) opt: 969
z-score: 1160.9 E(): 0; 99.3% identity in 142 aa

	overlap. Tbpase score is 0.897"
	/codon_start=1
	/transl_table=1
	/product="hypothetical protein RV044lc"
	/protein_id="CAAI7398.1"
	/db_xref="GI:2909516"
	/db_xref="SPTRMBL:Q50813"
	/translation="MGAKVDLKRLAALPDYPRAVLTIVDDGHRHRTVAVEPYLRREI PDEGDGPAAVVLDLGIRTRONLAAREVELTLPSPDSOSYLIDVGRAOASDAGPP DDTVARGCVPRVALHLRDAAPDSPTAAKGCGLHVDFSVF"
gene	complement(4650 . 6113)
	/gene="ppe"
CDS	complement(4650 . 6113) /note="RV0442c, (MTV037_06c), len: 487. Unknown member of ppe family, nearly identical to hypothetical protein from Mycobacterium tuberculosis (strain Erdman) and to protein fragment from Mycobacterium bovis, ANS S46909.1 (302 aa). FASTA scores: sp P42611 YHS_MCTU HYPOTHETICAL 50.6 KD PROTEIN IN H (517aa) opt: 3144 z-score: 244.9 E(): 0; 98.4% identity in492 aa overlap; and S46909IS46909.1 (302 aa) opt:1897 z-score: 1484.3 E(): 0; 98.0% identity in302 aa overlap. Tbpase score is 0.903"
	/codon_start=1
	/transl_table=1
	/product="ppe"
	/protein_id="CAAI7399.1"
	/db_xref="GI:2909517"
	/db_xref="SPTRMBL:Q53727"
	/translation="MTPSPHFMPPEITNSALMFAPDGSGPLIAATWANGELAEKLNS IASGSVTSEITSAMIGPSAMAANAAYOTVLNSTRADAEQAADAQAVATAFEHEE ALAATOPAVAANRRGMQLAATNWFOGNPALMDVEGALEOMNALDVAMAYHFEHEE ASAVAOLAQPQQVILRNGLDIGNGOINLFGRGTSCGINNNIGNNISCGSTGTGGC NIGSATNQMLGKLNLDGDNIFGFNTSGNIICFIGEDHDMGFGSSGNTGFGB SGTNWGFAFNSSGNIGTGNGSNLSGITGETTINGASGSLINTSFNMGMONAALI GSAAGEAFVYSAGAYATGKMSTAALSGLIIASLAGTGHOLANTNLNGLTFNPPI AAPRASVPYGILDSCNPNRGSSAAAAGSANACGLSPGISYFSFVNSSGNDGLNTAPA REPSTPGSIGPKSNFYPSPRESAVFASPTIQPVGSE"
gene	6295 . 6810
	/gene="RV0443"
CDS	6295 . 6810
	/gene="RV0443"
	/note="RV0443, (MTV037_07), len: 171. Unknown. Alternative start possible. Tbpase score is 0.915" /codon_start=1
	/transl_table=1
	/product="hypothetical protein RV0443"
	/protein_id="CAAI7400.1"
	/db_xref="GI:2909518"
	/db_xref="SPTRMBL:Q53728"
Query Match	43.9%; Score 709.8; DB 2; Length 19781;
Best Local Similarity	66.0%; Pared No. 5.5e-71:
Matches 1048; Conservative	0; Mismatches 532; Indels 9; Gaps 1;
Oy	3 caagctgatcgataatgcgacgaaccggcggttcgcccatgtagtgccgcatygacaagt 62
Dd	2512 CAAGCAATTTCGGTAGCAGAAGAGGCCCGTTCGGCCCTCGAGCGGGGCTTTGAACGCCCT 2571
Oy	63 ggcgcgaaccttggcggtttgagaacttggcgccggcgaggcaatttgtgttctggccaagc 122
Dd	2572 CGCCCATCCGGTAAAGGTGACATTG6GGCCCCCAAAGGCGCAACGTTCCTTGAAAAAGNA 2631
Oy	123 gttttggagaccaaggtttaaacgaaggcgttaacglttggacgttatgatatgagctga 182
Dd	2632 GTGGGGTGCCCCCACGATCACCAACAAGTGTGTGTTCATCCGCAAGAGATGACTGGA 2691
Oy	183 agatcgttttgaagacttggcgcccccaagcttgttgaagtcgttgcgcacaagaacacga 242
Dd	2692 GGATCCGTACGAGAAGATCGGGCGGAGCTGTGTCAAAAGAGTAAACCAGAACCCGATGA 2751
Oy	243 tqtlltcggttgaaggcacaccacgaacacatlctttgcgcgaagcattgatlaagggcgg 302

[illegible]

Db 1712 CTGAGCTGAGACTTCACAGAGGCGATGGCTTCGACAAAGGCTACTGCGCGTACTTC 1771
QY 604 gttaacgacttcgataaaccagcagcggtgctcgaagacggtatctctgtataccaa 663
Db 1772 GTAGAGACCAAGAGCGGATGGAGCCGCTCCAGAGACCCGTACATCCTGATCAACAG 1831
QY 664 gacaagatcagctcgtcccatctgttgcattgctgaaagtg-----tgcaaga 717
Db 1832 GGCAGAGATCTCTCATGGCGGACCTGCTGCGGCTGTGAGANAAGTTCATCCAGGCCAAC 1891
QY 718 acggtlaagccaactactatcgttgcgtgaagaagcgtggaagcgaagcgtltggaagcgtg 777
Db 1892 GCCTCAAGCCGCTCTGATCATTCGCCGAGACCTGAGGCGGAGCGGCTCTCCACCCCTC 1951
QY 778 gtctgcaacgagcttcgaagaagcgttgaagagctgcgaggtcaagggcgctacttggt 837
Db 1952 GTCTGTAACAGATCCGCGGACCTTCAACGCGGTGCGGCTCAAGGCCCGGCTTCGGC 2011
QY 838 gaccgcgtaagcgctccttgaagacctgagcggtgagcggtgccaagtggtcaac 897
2012 GACCCGCCCAAGCGCATGCTGAGAGACATGGCGTCTCACCGGCGCCACGGTCACTTCG 2071
QY 898 cccgagccgagcatltgtctgagcgagtgagcggtgaggtgctgagctcgccgagcg 957
Db 2072 GAGGAGGTGCGGCTCAAGCTGACACGATCGGCTCGAGGTGCTCGGACACCGCCGCGC 2131
QY 958 gtgtgtgtcagaagaagacagcagtgatctgtcagcgcgagcgagcagaaagcggtg 1017
Db 2132 ATCACCGTACCAAGAGGACACACAGATCTGACGGTCCGCGCAAGCGCGACGAGGTC 2191
QY 1018 gccaaaccgagcgaaaccacttgcgtcgaatcgaacaagaagcgatctgagtgatcg 1077
Db 2192 CAGGCGCCCATTCGCCACATCAAGGCCGAGATCGAAGAACGCGACTCCGACTGGGACCGC 2251
QY 1078 gaaaagcttgcgagcgagctgagcgaactgagcgagcggtgtgtgtatcaagtggt 1137
Db 2252 GAGAAGCTTCAAGAGACCGCTCGCAAGCTGCGCGGCGGTGTGCGTCAAGGTGCGC 2311
QY 1138 gccgcaacgagcaaccgcaactcaagaagcgcaagaagcgltcgaagatgctgagcg 1197
Db 2312 GCCGCCACCGAGGTGAGAGGCGCAAGACCGCTGAGAGACCGCATCTCCGCG 2371
QY 1198 gccaaagcgagctgcgaggaagcgatcgtccctgtgtgaggaagcctgcgtcatccacag 1257
Db 2372 ACCCGCGCGCGGTGAGAGAGGCGATCTCGGTGCTCGGCTCGGCTCGCACGCG 2431
QY 1258 gccgcaagcgagctgagcgaactgctgctgagcagcgltgacgaagctcctggtgc 1317
2432 GTCAAGGTGCTCGAAGGCAACCTCG-----CAAGACCGGCGAGGCGCACCGGTGC 2485
QY 1318 gacgtgtcttcgaagcccttgccgagcgctgtgtctgtatgcgcgccaagcgtgttg 1377
Db 2486 GCGGTCTCCGCGCGCGGTGAGCGCGCTCGCGTGTGATCGCGAGAACCGCGCGCTG 2545
QY 1378 gacgctcgtgtgtgtgtcaacaaggtcaagcgatcaccgcccgggcatlgygctgaagtg 1437
Db 2546 GAGGTTACGTATCACTCCAAAGTTCGCGACCTCGACACAAGGCGCCAGGCTTCAACGCC 2605
QY 1438 aacaacctgactatggtacttgcgctgagcggtcatcgaaccggttcaagtgact 1497
Db 2606 GCCACCGCGGAGTACGGAGACTGCTCAAGGCGCGGTATCGACCGGTGAAGGTCAAC 2665
QY 1498 aggttcgagcggtgtgaagcggtcatcgtgtgagcgatgtgtatcaaccgagaagcggtc 1557
Db 2666 CGCTCCGCGGTGAGACCGCGCTCATCGCTCTCTGTCAGCAGCAGACCGCCCTG 2725
QY 1558 gtgttcgaacagcgagcgaagaagatcaagacatcaccacgagcgagcgact 1615
Db 2726 GTCTGTAAGAAAGAGAGAGAGAGAGCGCGCGGTGCGCACAGCCACGCGCACT 2783

Search completed: January 12, 2001, 19:59:14
Job time: 26128 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 5, 2001, 14:25:08 ; Search time 115.27 Seconds
(without alignments)
547.045 Million cell updates/sec

Title: US-09-461-774-2
Perfect score: 2654
Sequence: 1 SKLEYDETARAMEVGMKX.....VVVDKRAKEDHHHGHAN 538

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues
otal number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: SP:REMBL_15:*
2: SP:archaea:*
3: SP:fungi:*
4: SP:human:*
5: SP_invertebrate:*
6: SP_mammal:*
7: SP_mhc:*
8: SP_organelle:*
9: SP_phage:*
10: SP_plant:*
11: SP_rodent:*
12: SP_virus:*
13: SP_vertebrate:*
14: SP_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1642	61.9	541	2	086801	086801 streptomyc
2	1580.5	59.6	541	2	09KXU5	09KXU5 streptomyc
3	1561.5	58.8	541	2	033659	033659 streptomyc
4	1557.5	58.7	541	2	09KGM1	09KGM1 mycobacteri
5	1525	57.5	539	2	P97086	P97086 tsukamurell
6	1509.5	56.9	544	2	09KZU4	09KZU4 propionibac
7	1427.5	53.8	539	2	09RC20	09RC20 bacillus sp
8	1426.5	53.7	543	2	09RAM4	09RAM4 thermus aqu
9	1423.5	53.6	544	2	09KFC3	09KFC3 bacillus ha
10	1422	53.6	538	2	09WYX6	09WYX6 thermotoga
11	1405.5	52.0	540	2	09KXCA9	09KXCA9 rhodotherm
12	1402.5	52.8	548	2	09RMQ9	09RMQ9 rhodococcus
13	1384	52.1	546	2	031198	031198 leptospira
14	1374.5	51.8	540	2	09KXRS	09KXRS streptococ
15	1372.5	51.7	544	2	032847	032847 lactobacill
16	1367.5	51.5	540	2	09KXFO	09KXFO clostridium
17	1365.5	51.5	543	2	057002	057002 synecchococ
18	1363.5	51.4	540	2	09KJVT	09KJVT clostridium
19	1355.5	51.1	548	2	087888	087888 lawsonia in

20	1345	50.7	542	2	09L691	09L691 rhizobium 1
21	1336	50.3	477	2	033658	033658 streptomyc
22	1322.5	49.8	543	2	09KJ23	09KJ23 lactobacill
23	1310.5	49.4	547	2	09PPR2	09PPR2 xyloella fas
24	1296	48.8	540	2	068324	068324 lactobacill
25	1293.5	48.7	540	2	09WML4	09WML4 methyllovoru
26	1292.5	48.7	538	2	033688	033688 sitophilus
27	1291.5	48.7	544	2	09J079	09J079 chlamydia p
28	1291	48.6	544	2	09L690	09L690 rhizobium 1
29	1289	48.6	546	2	09ZM50	09ZM50 heliocobacte
30	1288.5	48.5	544	2	09KMR7	09KMR7 vibrio chol
31	1278	48.2	550	2	09S303	09S303 rickettsia
32	1276.5	48.1	548	2	09L7P5	09L7P5 vibrio para
33	1273	48.0	547	2	09XAU7	09XAU7 alternomonas
34	1265	47.7	544	2	09KX73	09KX73 bacteroides
35	1258.5	47.4	555	2	09X603	09X603 primary end
36	1254.5	47.3	457	2	09K157	09K157 gardnerella
37	1251.5	47.2	548	2	008499	008499 rhopalosiph
38	1249.5	47.1	548	2	008500	008500 sitobion av
39	1238.5	46.7	497	2	09S6B0	09S6B0 chlamydia p
40	1237.5	46.6	497	2	09XAG5	09XAG5 chlamydophi
41	1230.5	46.4	503	2	087271	087271 bartoneella
42	1224.5	46.1	481	2	033733	033733 streptococ
43	1224.5	46.1	497	2	09X4G4	09X4G4 chlamydophi
44	1213.5	45.7	530	8	09TL21	09TL21 cyanidium c
45	1198.5	45.2	530	2	09KIC6	09KIC6 vibrio chol

ALIGNMENTS

RESULT 1
ID 086801 PRELIMINARY; PRT; 541 AA.
AC 086801;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE 60 KDA CHAPERONIN (PROTEIN CPN60) (GROEL PROTEIN) (HEAT SHOCK PROTEIN 60).
GN GROEL.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC SRRAIN-A3(2);
RA Saunders D., Harris D.,
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC SRRAIN-A3(2);
RA Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC SRRAIN-A3(2);
RA MEDLINE=97000351; PubMed=843436;
RA Redenbach M., Kleser H.M., Denapate D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
CC -1- FUNCTION: PREVENTS MISFOLDING AND PROMOTES THE REFOLDING AND
PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED UNDER STRESS
CONDITIONS (BY SIMILARITY)
CC -1- SUBUNIT: OLIGOMER OF 14 SUBUNITS COMPOSED OF TWO STACKED RINGS OF
7 SUBUNITS (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.
DR EMBL: AL031317; CAA20418.1; -;
DR HSP: P06139; UNON.
DR INTERPRO: IPR001844; -.

DR INTERPRO: IPR002423; -
DR PFAM: PF00118; cpn60_TCP1; 1.
DR PRINTS: PR00298; CHAPERONIN60.
DR PRINTS: PR00304; TCOMPLEXCP1.
DR PROSITE: PS00296; CHAPERONIN5_CPN60; 1.
KW Chaperone; ATP-binding.
SQ SEQUENCE 541 AA; 57118 MW; C2EFED580B33FFD8 CRC64;

Query Match 61.9%; Score 1642; DB 2: Length 541;
Best Local Similarity 60.3%; Pred. No. 1.9e-76;
Matches 327; Conservative 92; Mismatches 117; Indels 6; Gaps 3;

```
QY 1 SKLIYDETARRAMEVGMKDLADTVRVTLGPRGRHVVLAKAFGCPVTINDGVARETEL 60
DB 2 AKILKFEDARALERGVNKLADTVKTLGPRGRNVVIDKKGAPVTINDGVARETEV 61
QY 61 EDPFEDLGAOLVKSVAATKNDVAGDGTATVLAQALIKGRLVLAAGVNPVALGVGIG 120
DB 62 EDPYENLGAOLVKEVATKNDVAGDGTATVLAQALVREGKLVNVAAGASPALEKKGIDA 121
QY 121 AADAVEALLASAPFVSCKTGAQVATVSSRDEQIGDLVGEAMKNGHDSVSVKESSTL 180
DB 122 AVAAVSEDLATARPIDERSDIAAVALSAODQVGEILAEAMDVKGQVITVEESNTF 181
QY 181 GTELEFTEGIGFHKGFSLAYEVTDFDNOQAVLEDALELHODKISSLPDLPLEKVV--A 238
DB 182 GLELDTFGMAFDKYLSPFYTTDERMEAVLDPRYILLNCKISSINDLPLEKVIQA 241
QY 239 GTGKPLLIADVEGEALATLVNNAIRKTLAAVAVKGPFGDRRKAFLEDLAVVTGHHV 298
DB 242 NASKLLIITAEDELEAEALSTLVNKRIGTFNVAVAKAFGDRRKAMLDMAVLTGATVI 301
QY 299 NPDAIVAREVLEGLVGSARVVYVSKDPTVIYDGGTAEAVANRANHRAETDKSDSMD 358
DB 302 SEEGKLADQVLEGLVGTARITVTKDDTTVDGAKRDEVOGRTAQIKAELENTDSDMD 361
QY 359 REKLGELRLAKLAGVAATVKGATDTALKERKESVEDAVAAAKAVEGIVPGGASLIH 418
DB 362 REKLGELRLAKLAGVCVIVKGATVELEKRRHLEDAISATRAVEGIVSGGSALVH 421
QY 419 QARKALTELRLASLTGDEVLDVDFSEALAPLFWIAANAGLDGSSVNVKSELPAHGHLN 478
DB 422 AVK--VLEGNLKTGDEDETVGAVVARRAAVEPLRMIAENAGLEGIVYTKVADLDLDCGCFN 479
QY 479 VNTLSYGLAADGVIDPVKTRSAVLNASSVAWMLTTEVTVVDKPAKAE--HDHHGH 536
DB 480 AATGEGDLVAKAGVIDPVKTRSALENASTASLLTTEVLVEKKEDEEPAGSHSHG 539
QY 537 AH 538
DB 540 SH 541
```

RESULT 2
Q9KXU5 PRELIMINARY; PRT; 541 AA.

AC 09KXU5;
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, last sequence update)
DT 01-OCT-2000 (TREMblrel. 15, last annotation update)
DE CHAPERONIN 2.
GN GROEL2.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomycetes.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RA SRRAIN-A3(2); Harris D.;
RA Seeger K.J.; Harris D.;
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
EN [2]

RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Cerdeno A.M.; Parkhill J.; Barrell B.G.; Rajandream M.A.;
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Redenbach M.; Kleser H.M.; Denapalte D.; Eichner A.; Cullum J.;
RA Kinashl H.; Hopwood D.A.;
RT "A set of ordered cosmid and a detailed genetic and physical map for
the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL; AL357432; CAB93056.1; -
SQ SEQUENCE 541 AA; 56829 MW; 26BD7E89A9CB81C5 CRC64;

Query Match 59.6%; Score 1580.5; DB 2: Length 541; {
Best Local Similarity 60.9%; Pred. No. 2.6e-73;
Matches 321; Conservative 79; Mismatches 124; Indels 3; Gaps 2;

```
QY 1 SKLIYDETARRAMEVGMKDLADTVRVTLGPRGRHVVLAKAFGCPVTINDGVARETEL 60
DB 2 AKILKFEDARALERGVNKLADTVKTLGPRGRNVVIDKKGAPVTINDGVARETEV 61
QY 61 EDPFEDLGAOLVKSVAATKNDVAGDGTATVLAQALIKGRLVLAAGVNPVALGVGIG 120
DB 62 EDPYENLGAOLVKEVATKNDVAGDGTATVLAQALVREGKLVNVAAGASPALEKKGIDA 121
QY 121 AADAVEALLASAPFVSCKTGAQVATVSSRDEQIGDLVGEAMKNGHDSVSVKESSTL 180
DB 122 AVAAVSEDLATARPIDERSDIAAVALSAODQVGEILAEAMDVKGQVITVEESNTF 181
QY 181 GTELEFTEGIGFHKGFSLAYEVTDFDNOQAVLEDALELHODKISSLPDLPLEKVV--A 238
DB 182 GLELDTFGMAFDKYLSPFYTTDERMEAVLDPRYILLNCKISSINDLPLEKVIQA 241
QY 239 GTGKPLLIADVEGEALATLVNNAIRKTLAAVAVKGPFGDRRKAFLEDLAVVTGHHV 298
DB 242 NASKLLIITAEDELEAEALSTLVNKRIGTFNVAVAKAFGDRRKAMLDMAVLTGATVI 301
QY 299 NPDAIVAREVLEGLVGSARVVYVSKDPTVIYDGGTAEAVANRANHRAETDKSDSMD 358
DB 302 SEEGKLADQVLEGLVGTARITVTKDDTTVDGAKRDEVOGRTAQIKAELENTDSDMD 361
QY 362 REKLGELRLAKLAGVAATVKGATDTALKERKESVEDAVAAAKAVEGIVPGGASLIH 418
DB 362 REKLGELRLAKLAGVCVIVKGATVELEKRRHLEDAISATRAVEGIVSGGSALVH 421
QY 419 QARKALTELRLASLTGDEVLDVDFSEALAPLFWIAANAGLDGSSVNVKSELPAHGHLN 478
DB 422 AVK--VLEGNLKTGDEDETVGAVVARRAAVEPLRMIAENAGLEGIVYTKVADLDLDCGCFN 479
QY 479 VNTLSYGLAADGVIDPVKTRSAVLNASSVAWMLTTEVTVVDKPAKAE--HDHHGH 536
DB 480 AATGEGDLVAKAGVIDPVKTRSALENASTASLLTTEVLVEKKEDEEPAGSHSHG 539
QY 537 AH 538
DB 540 SH 541
```

RESULT 3
O33659 PRELIMINARY; PRT; 541 AA.

AC 033659;
DT 01-JAN-1998 (TREMblrel. 05, Created)
DT 01-JAN-1998 (TREMblrel. 05, last sequence update)
DT 01-MAY-2000 (TREMblrel. 13, last annotation update)
DE 60 KDA CHAPERONIN (PROTEIN CPN60) (GROEL PROTEIN) (HEAT SHOCK PROTEIN 60).
GN GROEL1.
OS Streptomyces lividans.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomycetes.
OX NCBI_TaxID=1916;
RN [1]
RP SEQUENCE FROM N.A.

RP SEQUENCE FROM N.A.
RC STRAIN-ITI 376;
RA Thorofsdottir E.T.T., Backman V.M., Blondal T.,
RA Thorbjarnardottir S.H., Palsdottir A., Hauksdottir H.,
RA Kristjansdottir S., Eggertsson G.,
RT "Heat shock in Rhodothermus marinus: Cloning and sequence analysis of
RT the groEL, dnaK and dnaJ genes."
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
CC - FUNCTION: PREVENTS MISFOLDING AND PROMOTES THE REFOLDING AND
CC PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED UNDER STRESS
CC CONDITIONS (BY SIMILARITY).
CC - SUBUNIT: OLIGOMER OF 14 SUBUNITS COMPOSED OF TWO STACKED RINGS OF
CC 7 SUBUNITS (BY SIMILARITY).
CC - SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.
DR EMBL: AF145252; AAD37976.1; -.
DR HSSP: P06139; IGRU.
DR INTERPRO: IPR001844; -.
DR INTERPRO: IPR002423; -.
DR PFAM: PF00118; cpn60_TCP1; 1.
DR PRINTS: PR00298; CHAPERONIN60.
DR PRINTS: PR00304; TCOMPLEXTCP1.
DR PROSITE: PS00296; CHAPERONINS_CPN60; 1.
KW Heat shock; Chaperone; ATP-binding.
SQ SEQUENCE 540 AA; 57666 MW; 6867448B7BEF18EC CRC64;

Query Match 53.0%; Score 1405.5; DB 2; Length 540;
Best Local Similarity 53.2%; Pred. No. 2,3e-64;
Matches 291; Conservative 87; Mismatches 146; Indels 3; Gaps 3;

QY 1 SKLIEDETRARRAMEYMDKADTVRTVLTGPRGRHVYLAKEGPGPTVNDGVTARETEL 60
DB 3 AKQITFNADARMALKRGVCKLADAVKVTTLGPKGRNVIIEKFPAPYTKRGVYAKTEL 62
QY 61 EDDPEDGAGLVKSVATKTNVAGDGTATTATLAAQLIKGRLVAGVNPVALGVGICK 120
DB 63 EDKLENGAGMVEKVASKTSDVAGDGTATTATLAAQLIKGRLVAGVNPVALGVGICK 122
QY 121 AANAVERALLASATPPVSGKGTGIAQVATVSSR-DEQIGDLVGEAMNKVGHDSVSKESST 179
DB 123 AVEVVVAELKKMSQEVODKRRIAQVATISANGKAIGQLIADAFKVGKGVITVEBAGK 182
QY 180 LGTELEETEGIGFHKGLSAFVTFDPONOQAVLEDAIILHODKISSLPDLLPLEKAVG 239
DB 183 TETLEVEVEGMDRGRLSYETNPDMFAVLEDAIILHDKKISMKRLPLEKAVG 242
QY 240 TGRPLLIADVEGEALATLVNAIRKTLKAVAVKGPFGDRRAKAFLEDAVVTGGHYVN 299
DB 243 TGRPLLIADVEGEALATLVNKLKGLVLAVAVKAPGFGDRRAKAFLEDAIILTGTVIS 302
QY 300 PDAGIVLREGLVGLSARRVYVSKDDTVYVDGGGTRAEVANRANHRAEIDKSDSDMR 359
DB 303 EEKGYRLNATLDLYGQAEKRIYDKDNTTIVGKGPQAKARANOIRQOIEETTSYDR 362
QY 360 EKIAGERLAKLAGVAIVKVGATDTALKEKRESVEDAVAAAKAAVEGIVPGGASLIHQ 419
DB 363 EKIQERLAKLAGVAIVKIGATTEPEKKEKARVEDALHTRAAVEGIVPGGAYI-R 421
QY 420 ARKALTELRLASLTGDEVLYGVDFSEALAPLFWIAANAGLDGSVVYVKNVSELPAHGILNV 479
DB 422 AIALADKVEVE-NEDQKIGVQIVORALIEPLRLQIAANAGGESSIVYQVRVKEGGCGFN 480
QY 480 NTLSTYGDLAADGVIDPVKVTRSVAVLNASSVARNVLTTEVVVDKPAK 526
DB 481 QTEEFGLTEOGVIDPTKARTALENNAASYAGLLTTEAVVAEKPER 527

RESULT 12
Q9RM09 PRELIMINARY; PRT; 548 AA.
AC Q9RM09;
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)

DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)
DE GROEL PROTEIN.
GN DR0607.
OS Deinococcus radiodurans.
OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-R1;
RX MEDLINE=20036896; PubMed=10567266;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwin M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Uitterback T., Zalewski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome Sequence of the Radioresistant Bacterium Deinococcus
RT radiodurans R1."
RT Science 286:1571-1577(1999).
DR EMBL: AE001918; AAF10186.1; -.
DR HSSP: P45746; ISRV.
DR TIGR: DR0607; -.
DR INTERPRO: IPR001844; -.
DR INTERPRO: IPR002423; -.
DR PFAM: PF00118; cpn60_TCP1; 1.
DR PRINTS: PR00296; CHAPERONIN60.
DR PRINTS: PR00304; TCOMPLEXTCP1.
SQ SEQUENCE 548 AA; 57777 MW; EC8829C370F6675D CRC64;

Query Match 52.8%; Score 1402.5; DB 2; Length 548;
Best Local Similarity 53.9%; Pred. No. 3,4e-64;
Matches 285; Conservative 96; Mismatches 145; Indels 3; Gaps 3;

QY 1 SKLIEDETRARRAMEYMDKADTVRTVLTGPRGRHVYLAKEGPGPTVNDGVTARETEL 60
DB 2 AKQVFTDESARRSLERGVANAVANAVKVTTLGPKGRNVIIEKFPAPYTKRGVYAKTEL 61
QY 61 EDDPEDGAGLVKSVATKTNVAGDGTATTATLAAQLIKGRLVAGVNPVALGVGICK 120
DB 62 EDKLENGAGLVKSVATKTNVAGDGTATTATLAAQLIKGRLVAGVNPVALGVGICK 121
QY 121 AANAVERALLASATPPVSGKGTGIAQVATVSSRDEQIGDLVGEAMNKVGHDSVSKESST 180
DB 122 AANAVERALLASATPPVSGKGTGIAQVATVSSRDEQIGDLVGEAMNKVGHDSVSKESST 181
QY 181 GTELEETEGIGFHKGLSAFVTFDPONOQAVLEDAIILHODKISSLPDLLPLEKAVG 240
DB 182 DTEVDVVEGMDRGRLSYETNPDMFAVLEDAIILHDKKISMKRLPLEKAVG 241
QY 241 GTELEETEGIGFHKGLSAFVTFDPONOQAVLEDAIILHODKISSLPDLLPLEKAVG 240
DB 242 GTELEETEGIGFHKGLSAFVTFDPONOQAVLEDAIILHODKISSLPDLLPLEKAVG 241
QY 241 GTELEETEGIGFHKGLSAFVTFDPONOQAVLEDAIILHODKISSLPDLLPLEKAVG 240
DB 242 GTELEETEGIGFHKGLSAFVTFDPONOQAVLEDAIILHODKISSLPDLLPLEKAVG 241
QY 301 DAGIVLREGLVGLSARRVYVSKDDTVYVDGGGTRAEVANRANHRAEIDKSDSDMR 360
DB 302 DAGIVLREGLVGLSARRVYVSKDDTVYVDGGGTRAEVANRANHRAEIDKSDSDMR 361
QY 361 KIGERLAKLAGVAIVKVGATDTALKEKRESVEDAVAAAKAAVEGIVPGGASLIHQ 420
DB 362 KIGERLAKLAGVAIVKVGATDTALKEKRESVEDAVAAAKAAVEGIVPGGASLIHQ 420
QY 421 RKALTELRLASLTGDEVLYGVDFSEALAPLFWIAANAGLDGSVVYVKNVSELPAHGILNV 479
DB 422 IPAYRKAASLTGDEATGARILIRALEPAPARQIAANAGGESSIVYVYVYVYVYVYVYV 480
QY 480 NTLSTYGDLAADGVIDPVKVTRSVAVLNASSVARNVLTTEVVVDKPAK 527
DB 481 ATGEYEDVAVAGIVDPKARTALONNASIGALITTEAVVADKPEKA 529

RESULT 13

031198
ID 031198 PRELIMINARY; PRT: 546 AA.
AC 031198:
DT 01-JAN-1998 (Tremblrel. 05, Created)
DT 01-JAN-1998 (Tremblrel. 05, last annotation update)
DT 01-MAY-2000 (Tremblrel. 13, last annotation update)
DE 60 KDA CHAPERONIN (PROTEIN CPN60) (GROEL PROTEIN) (HEAT SHOCK PROTEIN 60).
GN HSP58.
OS Leptospira interrogans.
OC Bacteria; Spirochaetales; Leptospiraceae; Leptospira.
OX NCBI_TaxID=173;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HY-1;
RA Kim M.J., Ahn B.Y.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: PREVENTS MISFOLDING AND PROMOTES THE REFOLDING AND PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED UNDER STRESS CONDITIONS (BY SIMILARITY).
CC -1- SUBUNIT: OLIGOMER OF 14 SUBUNITS COMPOSED OF TWO STACKED RINGS OF 7 SUBUNITS (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.
DR EMBL: AF032910; AAB86965.1; -.
DR HSSP: P06139; 1GRL.
DR INTERPRO: IPR001844; -.
DR INTERPRO: IPR002423; -.
DR PFAM: PF00118; cpn60.TCPI.1.
DR PRINTS: PR00298; CHAPERONIN60.
DR PRINTS: PR00304; TCOMPLEXCPI.
DR PROSITE: PS00296; CHAPERONINS.CPN60; 1.
DR Heat shock; Chaperone; ATP-binding.
KW SEQUENCE 546 AA; 57969 MW; 15BC9CE0A8E2A6E3 CRC64;

Query Match 52.1%; Score 1384; DB 2; Length 546;
Best Local Similarity 52.2%; Pred. No. 2.9e-63;
Matches 275; Conservative 106; Mismatches 142; Indels 4; Gaps 3;

031198
ID 031198 PRELIMINARY; PRT: 546 AA.
AC 031198:
DT 01-JAN-1998 (Tremblrel. 05, Created)
DT 01-JAN-1998 (Tremblrel. 05, last annotation update)
DT 01-MAY-2000 (Tremblrel. 13, last annotation update)
DE 60 KDA CHAPERONIN (PROTEIN CPN60) (GROEL PROTEIN) (HEAT SHOCK PROTEIN 60).
GN HSP58.
OS Leptospira interrogans.
OC Bacteria; Spirochaetales; Leptospiraceae; Leptospira.
OX NCBI_TaxID=173;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HY-1;
RA Kim M.J., Ahn B.Y.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: PREVENTS MISFOLDING AND PROMOTES THE REFOLDING AND PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED UNDER STRESS CONDITIONS (BY SIMILARITY).
CC -1- SUBUNIT: OLIGOMER OF 14 SUBUNITS COMPOSED OF TWO STACKED RINGS OF 7 SUBUNITS (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.
DR EMBL: AF032910; AAB86965.1; -.
DR HSSP: P06139; 1GRL.
DR INTERPRO: IPR001844; -.
DR INTERPRO: IPR002423; -.
DR PFAM: PF00118; cpn60.TCPI.1.
DR PRINTS: PR00298; CHAPERONIN60.
DR PRINTS: PR00304; TCOMPLEXCPI.
DR PROSITE: PS00296; CHAPERONINS.CPN60; 1.
DR Heat shock; Chaperone; ATP-binding.
KW SEQUENCE 546 AA; 57969 MW; 15BC9CE0A8E2A6E3 CRC64;

Db 479 LTMWEDMIOAGVVDPAKVBSALONASIGSMILTTEVTITDKPKD 525

RESULT 14
ID 09X4R5 PRELIMINARY; PRT: 540 AA.
AC 09X4R5:
DT 01-NOV-1999 (Tremblrel. 12, Created)
DT 01-NOV-1999 (Tremblrel. 12, last sequence update)
DT 01-OCT-2000 (Tremblrel. 15, last annotation update)
DE 60 KDA CHAPERONIN (PROTEIN CPN60) (GROEL PROTEIN) (HEAT SHOCK PROTEIN 60).
GN GROEL.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CP1200;
RA Kim S.N., Lee J.Y., Kim S.W., Choi I.H., Rhee D.K.;
RT "GROEL sequences in Streptococcus pneumoniae."
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: PREVENTS MISFOLDING AND PROMOTES THE REFOLDING AND PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED UNDER STRESS CONDITIONS (BY SIMILARITY).
CC -1- SUBUNIT: OLIGOMER OF 14 SUBUNITS COMPOSED OF TWO STACKED RINGS OF 7 SUBUNITS (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.
DR EMBL: AF117741; AAD23455.1; -.
DR HSSP: P06139; 1JON.
DR INTERPRO: IPR001844; -.
DR INTERPRO: IPR002423; -.
DR PFAM: PF00118; cpn60.TCPI.1.
DR PRINTS: PR00298; CHAPERONIN60.
DR PRINTS: PR00304; TCOMPLEXCPI.
DR PROSITE: PS00296; CHAPERONINS.CPN60; 1.
DR Chaperone; ATP-binding.
KW SEQUENCE 540 AA; 57213 MW; 4CDB489AF1C632DA CRC64;

Query Match 51.8%; Score 1374.5; DB 2; Length 540;
Best Local Similarity 53.1%; Pred. No. 8.8e-63;
Matches 278; Conservative 98; Mismatches 145; Indels 3; Gaps 1;

031198
ID 031198 PRELIMINARY; PRT: 546 AA.
AC 031198:
DT 01-JAN-1998 (Tremblrel. 05, Created)
DT 01-JAN-1998 (Tremblrel. 05, last annotation update)
DT 01-MAY-2000 (Tremblrel. 13, last annotation update)
DE 60 KDA CHAPERONIN (PROTEIN CPN60) (GROEL PROTEIN) (HEAT SHOCK PROTEIN 60).
GN HSP58.
OS Leptospira interrogans.
OC Bacteria; Spirochaetales; Leptospiraceae; Leptospira.
OX NCBI_TaxID=173;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HY-1;
RA Kim M.J., Ahn B.Y.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: PREVENTS MISFOLDING AND PROMOTES THE REFOLDING AND PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED UNDER STRESS CONDITIONS (BY SIMILARITY).
CC -1- SUBUNIT: OLIGOMER OF 14 SUBUNITS COMPOSED OF TWO STACKED RINGS OF 7 SUBUNITS (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.
DR EMBL: AF032910; AAB86965.1; -.
DR HSSP: P06139; 1GRL.
DR INTERPRO: IPR001844; -.
DR INTERPRO: IPR002423; -.
DR PFAM: PF00118; cpn60.TCPI.1.
DR PRINTS: PR00298; CHAPERONIN60.
DR PRINTS: PR00304; TCOMPLEXCPI.
DR PROSITE: PS00296; CHAPERONINS.CPN60; 1.
DR Heat shock; Chaperone; ATP-binding.
KW SEQUENCE 546 AA; 57969 MW; 15BC9CE0A8E2A6E3 CRC64;

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 5, 2001, 14:25:29 ; Search time 35.06 Seconds
(without alignments)
495,557 Million cell updates/sec

Title: US-09-461-774-2

Perfect score: 2654
Sequence: 1 SKLIEYDPTARAMEYGMK.....VYDKPAKEDHHHGHAAH 538

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 88757 seqs, 32294092 residues

Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2602	98.0	539	1	CH61_MYCTU
2	2223	83.8	537	1	CH61_MYCLE
3	1643.5	61.9	539	1	CH61_STRAL
4	1620.5	61.1	539	1	CH61_STRRCO
5	1574.5	59.3	539	1	CH62_MYCTU
6	1543.5	58.2	539	1	CH62_STRAL
7	1543.5	58.2	540	1	CH60_MYCPA
8	1543.5	58.2	540	1	CH62_MYCLE
9	1454.5	54.8	540	1	CH60_CLOTM
10	1440	54.3	540	1	CH60_THERBR
11	1426.5	53.7	542	1	CH60_THERH
12	1415.5	53.3	539	1	CH60_BACST
13	1401	52.8	543	1	CH60_CLOAB
14	1399	52.7	539	1	CH60_CLOPE
15	1398.5	52.7	540	1	CH60_BACSU
16	1393	52.5	540	1	CH61_SYN3
17	1393	52.5	544	1	CH61_STNVU
18	1392.5	52.2	538	1	CH60_BACP3
19	1385	52.2	546	1	CH60_LEPIN
20	1384.5	52.2	544	1	CH60_BACHD
21	1369	51.6	541	1	CH62_SYN3
22	1359	51.2	541	1	CH63_RHIME
23	1344	50.6	544	1	CH60_STNP7
24	1344	50.6	546	1	CH61_RHOSH
25	1339.5	50.5	538	1	CH60_STRAEP
26	1338	50.4	545	1	CH60_RHOCA
27	1337	50.4	552	1	CH60_COXBU
28	1335	50.3	539	1	CH61_BRAJA
29	1332.5	50.2	545	1	CH61_RHIME
30	1329.5	50.1	542	1	CH60_LACLA
31	1322.5	49.8	546	1	CH60_RHLIV
32	1321.5	49.8	547	1	CH60_BORPE
33	1320.5	49.8	546	1	CH60_BRUAB

34	1318.5	49.7	544	1	CH60_NEIMA	P57006 neisseria m
35	1318.5	49.7	546	1	CH60_CHRVI	P31293 chromatium
36	1318	49.7	538	1	CH60_STRAU	008854 staphylococ
37	1317.5	49.6	544	1	CH60_NEIMB	P42385 neisseria m
38	1316.5	49.6	545	1	CH60_AQUAE	067943 aquifex aeo
39	1315.5	49.6	544	1	CH60_AGRU	P30779 aquibacteri
40	1314.5	49.5	540	1	CH60_ERWCA	066220 erwina car
41	1314.5	49.5	543	1	CH60_BARBA	P35635 bartonella
42	1314.5	49.5	547	1	CH60_BARHE	033963 bartonella
43	1313.5	49.5	546	1	CH60_BURCE	092160 burkholderi
44	1312.5	49.5	540	1	CH60_ENTAE	066198 enterobacte
45	1312	49.4	542	1	CH62_RHIME	P35470 rhizobium m

ALIGNMENTS

```

RESULT 1
ID CH61_MYCTU STANDARD; PRT: 539 AA.
AC Q59573; Q59581;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE 60 KDA CHAPERONIN 1 (PROTEIN CPN60-1) (GROEL PROTEIN 1).
GN GROEL1 OR GROEL-1 OR RV3417C OR MTCY78.12.
OS Mycobacterium tuberculosis.
OC Bacteria: Firmicutes: Actinobacteria: Actinobacteridae;
OC Actinomycetales: Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squires R., Sulston J.E.,
RA Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ERDMANN;
RX MEDLINE=93219332; PubMed=7681982;
RA Kong T.H., Coates A.R.M., Butcher P.D., Hickman C.J., Shinnick T.M.;
RT "Mycobacterium tuberculosis expresses two chaperonin-60 homologs."
RL Proc. Natl. Acad. Sci. U.S.A. 90:2608-2612(1993).
CC - FUNCTION: PREVENTS MISFOLDING AND PROMOTES THE REFOLDING AND
CC PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED UNDER STRESS
CC CONDITIONS. (BY SIMILARITY).
CC - SUBUNIT: OLIGOMER OF 14 SUBUNITS COMPOSED OF TWO STACKED RINGS OF
CC 7 SUBUNITS (BY SIMILARITY).
CC - SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/annouce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: 277165; CAB01006.1; -.
CC EMBL: X60350; CAA42909.1; -.
CC HSP: P06139; IGR.
CC TUBERCULIST; RV3417C; -.
CC INTERPRO: IPR001844; -.
CC INTERPRO: IPR002423; -.
CC PFM: PF00118; cpn60_TCPI; 1.

```

DR PRINTS; PR00298; CHAPERONIN60.
 DR PRINTS; PR00304; TCOMPLEXCP1.
 DR PROSITE; PS00296; CHAPERONINS.CPN60; 1.
 KW Chaperone; ATP-binding; Multigene family; Antigen; Cell wall.
 FT VARIANT 467 467 N->K (IN STRAIN ERMANN).
 SQ SEQUENCE 539 AA; 55877 MW; 3E0B93164C091B63 CRC64;

Query Match 98.0%; Score 2602; DB 1; Length 539;
 Best Local Similarity 98.1%; Pred. No. 1.4e-117;
 Matches 528; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 1 SKLIEYDETARRAMEVGMKDLADTVRVTLGPRGRHVVLAKAFGGPTVNDGVYAREIEL 60
 DB 2 SKLIEYDETARRAMEVGMKDLADTVRVTLGPRGRHVVLAKAFGGPTVNDGVYAREIEL 61
 QY 61 EDPEFDLGAOLVKSVAITNDVAGDGTATTLAOLIKGGRILVAAGVNPALGVGIGK 120
 DB 62 EDPEFDLGAOLVKSVAITNDVAGDGTATTLAOLIKGGRILVAAGVNPALGVGIGK 121
 QY 121 AADAVEALLASAPVSGKTGIAQVATVSSRDEQIGDLVGEAMNKVGHDSVSVKESSTL 180
 DB 122 AADAVEALLASAPVSGKTGIAQVATVSSRDEQIGDLVGEAMNKVGHDSVSVKESSTL 181
 QY 181 GTELEFTEGIGFHKGFLSAYFVTFDNOQAVLEDAILLHODKISSLPDLLPLEKAVAGT 240
 DB 182 GTELEFTEGIGFHKGFLSAYFVTFDNOQAVLEDAILLHODKISSLPDLLPLEKAVAGT 241
 QY 241 GKPLIIVAEDEGEALATLVNAIKTKLAKAVKGFYFGDRKALLEDLAVVTGGQVNP 300
 DB 242 GKPLIIVAEDEGEALATLVNAIKTKLAKAVKGFYFGDRKALLEDLAVVTGGQVNP 301
 QY 301 DAGILREVGEVLGSARRVYVSKDVIYVGGGTAEAANRANHLRAEIKSDSDMDRE 360
 DB 302 DAGILREVGEVLGSARRVYVSKDVIYVGGGTAEAANRANHLRAEIKSDSDMDRE 361
 QY 361 KLGERLAKLAGVAVIKGAATDTALKERKESVEDAVAAKAAVEEGIVPGGASLIHQ 420
 DB 362 KLGERLAKLAGVAVIKGAATDTALKERKESVEDAVAAKAAVEEGIVPGGASLIHQ 421
 QY 421 RKALTELASITGDEVLDVDFSEALAPLFWIANAGLDGSVYVNVKSELPAHGGLVNP 480
 DB 422 RKALTELASITGDEVLDVDFSEALAPLFWIANAGLDGSVYVNVKSELPAHGGLVNP 481
 QY 481 TLSYGDLAADVIVPVKTRSAVLNASSVARMVLTTEVVVDKPAKADHDHNGHAN 538
 DB 482 TLSYGDLAADVIVPVKTRSAVLNASSVARMVLTTEVVVDKPAKADHDHNGHAN 539

RESULT 2
 CH61_MYCLE STANDARD; PRT: 537 AA.

ID CH61_MYCLE
 AC P37578;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 01-FEB-1996 (Rel. 33, Last annotation update)
 DE 60 KDA CHAPERONIN 1 (PROTEIN CPN60 1) (GROEL PROTEIN 1).
 GN GROEL-1 OR GROEL OR B229_C3_248 OR B1620_C3_228.
 OS Mycobacterium leprae.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-92374850; PubMed-1354834;
 RA de Wit T.F.R., Bekelie S., Ostland A., Miko T.L., Hermans P.W.M.,
 van Soelingen D., Drijfhout J., Schoeningh R., Janson A.A.M.,
 Thole J.E.R.;
 RT "Mycobacteria contain two groEL genes: the second Mycobacterium
 leprae groEL gene is arranged in an operon with groES.";
 RL Mol. Microbiol. 6:1995-2007 (1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Smith D.R., Robison K.;

RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
 CC -I- FUNCTION: PREVENTS MISFOLDING AND PROMOTES THE REFOLDING AND
 CC PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED UNDER STRESS
 CC CONDITIONS (BY SIMILARITY).
 CC -I- SUBUNIT: OLIGOMER OF 14 SUBUNITS COMPOSED OF TWO STACKED RINGS OF
 CC 7 SUBUNITS (BY SIMILARITY).
 CC -I- SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; Z11665; -; NOT_ANNOTATED_CDS.
 DR EMBL; U00015; AAC43228.1; -;
 DR EMBL; U00020; AAA17312.1; -;
 DR PIR; S25181; S25181.
 DR HSSP; P06139; IGRL.
 DR INTERPRO; IPR001844; -;
 DR INTERPRO; IPR002423; -;
 DR PFAM; PF00118; cpn60-TCPL; 1.
 DR PRINTS; PR00298; CHAPERONIN60.
 DR PRINTS; PR00304; TCOMPLEXCP1.
 DR PROSITE; PS00296; CHAPERONINS.CPN60; 1.
 KW Chaperone; Antigen; Multigene family.
 SQ SEQUENCE 537 AA; 55816 MW; F2355B1BCFADF50F CRC64;

Query Match 83.8%; Score 2223; DB 1; Length 537;
 Best Local Similarity 82.3%; Pred. No. 1.5e-99;
 Matches 443; Conservative 48; Mismatches 45; Indels 2; Gaps 2;

QY 1 SKLIEYDETARRAMEVGMKDLADTVRVTLGPRGRHVVLAKAFGGPTVNDGVYAREIEL 60
 DB 2 SKLIEYDETARRAMEVGMKDLADTVRVTLGPRGRHVVLAKAFGGPTVNDGVYAREIEL 61
 QY 61 EDPEFDLGAOLVKSVAITNDVAGDGTATTLAOLIKGGRILVAAGVNPALGVGIGK 120
 DB 62 EDPEFDLGAOLVKSVAITNDVAGDGTATTLAOLIKGGRILVAAGVNPALGVGIGK 121
 QY 121 AADAVEALLASAPVSGKTGIAQVATVSSRDEQIGDLVGEAMNKVGHDSVSVKESSTL 180
 DB 122 AADAVEALLASAPVSGKTGIAQVATVSSRDEQIGDLVGEAMNKVGHDSVSVKESSTL 181
 QY 181 GTELEFTEGIGFHKGFLSAYFVTFDNOQAVLEDAILLHODKISSLPDLLPLEKAVAGT 240
 DB 182 GTELEFTEGIGFHKGFLSAYFVTFDNOQAVLEDAILLHODKISSLPDLLPLEKAVAGT 241
 QY 241 GKPLIIVAEDEGEALATLVNAIKTKLAKAVKGFYFGDRKALLEDLAVVTGGQVNP 300
 DB 242 GKPLIIVAEDEGEALATLVNAIKTKLAKAVKGFYFGDRKALLEDLAVVTGGQVNP 301
 QY 301 DAGILREVGEVLGSARRVYVSKDVIYVGGGTAEAANRANHLRAEIKSDSDMDRE 360
 DB 302 DAGILREVGEVLGSARRVYVSKDVIYVGGGTAEAANRANHLRAEIKSDSDMDRE 361
 QY 361 KLGERLAKLAGVAVIKGAATDTALKERKESVEDAVAAKAAVEEGIVPGGASLIHQ 420
 DB 362 KLGERLAKLAGVAVIKGAATDTALKERKESVEDAVAAKAAVEEGIVPGGASLIHQ 420
 QY 421 RKALTELASITGDEVLDVDFSEALAPLFWIANAGLDGSVYVNVKSELPAHGGLVNP 480
 DB 422 RKALTELASITGDEVLDVDFSEALAPLFWIANAGLDGSVYVNVKSELPAHGGLVNP 480
 QY 481 TLSYGDLAADVIVPVKTRSAVLNASSVARMVLTTEVVVDKPAKADHDHNGHAN 538
 DB 482 TLSYGDLAADVIVPVKTRSAVLNASSVARMVLTTEVVVDKPAKADHDHNGHAN 537

RESULT 3

QY 481 TLSYGDLAADVDPVKYTRSAVLNASSVARNVLTTEVVDKPAKA 527
 DB 478 TGVYDMLAEGILDPKAYTRSAVLNASSVARNVLTTEVVDKPEKA 524

RESULT 7

ID CH60_MYCPE STANDARD: PRT: 540 AA.

AC P42384;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE 60 KDA CHAPERONIN (PROTEIN CPN60) (GROEL PROTEIN) (65 KDA ANTIGEN)
 DE (HEAT SHOCK PROTEIN 65).
 GN MOA OR GROEL OR HSP65.
 OS Mycobacterium paratuberculosis.
 CC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 CC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC Bujdoso R.;
 RL Submitted (AUG-1993) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=LINDA;
 RX MEDLINE=96157365; PubMed=8574825;
 RA El-Zaateri F.A.K., Naser S.A., Engstrand L., Burch P.E.,
 RA Hachem C.Y., Whipple D.L., Graham D.Y.;
 RT "Nucleotide sequence analysis and seroreactivities of the 65K heat
 RT shock protein from Mycobacterium paratuberculosis.";
 RL Clin. Diagn. Lab. Immunol. 2:657-664(1995).
 CC -1- FUNCTION: PREVENTS MISFOLDING AND PROMOTES THE REFOLDING AND
 CC PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED UNDER STRESS
 CC CONDITIONS (BY SIMILARITY).
 CC -1- SUBUNIT: OLIGOMER OF 14 SUBUNITS COMPOSED OF TWO STACKED RINGS OF
 CC 7 SUBUNITS (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch)

CC EMBL; X74518; CAAS2630.1; -;
 CC EMBL; U15989; AAA99670.1; ALT_INIT.
 CC HSSP; P06139; 1GRL.
 CC INTERPRO: IPR001844; -;
 CC INTERPRO: IPR002423; -;
 CC PFAM: PF00118; CPN60_TPCP1.1.
 CC PRINTS; PR00298; CHAPERONIN60.
 CC PRINTS; PR00304; TCOMPLEXTCP1.
 CC PROSITE; PS00296; CHAPERONINS_CPN60; 1.
 CC KMW Chaperone; ATP-binding.
 CC INT MET 0
 CC FT VARIANT 44 BY SIMILARITY.
 CC FT VARIANT 44 A -> S (IN STRAIN LINDA).
 CC FT VARIANT 428 P -> L (IN STRAIN LINDA).
 CC FT VARIANT 480 E -> K (IN STRAIN LINDA).
 CC FT VARIANT 490 AD -> TE (IN STRAIN LINDA).
 CC FT VARIANT 507 A -> S (IN STRAIN LINDA).
 CC FT VARIANT 524 A -> T (IN STRAIN LINDA).
 CC FT VARIANT 526 A -> P (IN STRAIN LINDA).
 CC SO SEQUENCE 540 AA; 36580 MM; 1B524C87CAFB10D8 CRC64;

Query Match 58.2%; Score 1543.5; DB 1; Length 540;
 Best Local Similarity 60.0%; Pred. No. 3.6e-67;
 Matches 316; Conservative 78; Mismatches 130; Indels 3; Gaps 2;

QY 1 SKLIEYDEARRGLERGINALADAVKVTLPGRNVVLEKKMGAPTITNDGVSIAKETEL 60

DB 1 AKTIAYDEARRGLERGINALADAVKVTLPGRNVVLEKKMGAPTITNDGVSIAKETEL 60
 QY 61 EDPEDDLGAOLVKYATNTNVDGTTATITLQAALIKGLRLVYAGVNVVAGVGIGK 120
 DB 61 EDPYKIEALVEKAYKATDVAGDGTATVAVLQALVREGLRVNAQAANPLGKRGIEK 120
 QY 121 AADAVEALLASATFVSGKTGLAQAVATSSRDEQIGDLVGCAMNKVGHDSVYKESSTL 180
 DB 121 AVEKTEITLLSAKEVEKDDQIAATAISAGDOSIGDLIAEMKVGEGVITVESENTEF 180
 QY 181 GTELEFTEGIFGHKGLSAYVVTDFDNOQAVLEDAJILLHODKISSDPLPLEKVAAGT 240
 DB 181 GLQLELTFEGMFPDQKXISGYEFTDAERQEALEDPFILLVSSKSTVADLLPLEKVIQA 240
 QY 241 GKPLLIIVADVEGEALATLVNNAIRKTIKAVAVGPGYGRDRKAFLELAVVTVGHVNP 300
 DB 241 GKPLLIIVADVEGEALATLVNNAIRKTIKAVAVGPGYGRDRKAFLELAVVTVGHVNP 300
 QY 301 DAGIVLEVGLEVLGSARRVVSCKDVIYDGGTGAEAANRANHLRAIDKSDMDRE 360
 DB 301 EYGLSLESADISLIGKARKVYVTDEITIVGAGDSDAIAGVQAQIRTEIENSDDYRE 360
 QY 361 KLGERLAKLAGVAVIKYGAATDTALKERKESVEDAVAAKAAVEGIVPGGASLIHQ 420
 DB 361 KLOERLAKLAGVAVIKYGAATEVELKERKRIEDAVRNKAAVEGIVAGGVALLH-A 419
 QY 421 RKALTELASLTGDEVLDVDFSEALAPLEWIAANAGLDSSVYVKNKSELPAGHGLNVN 480
 DB 420 IPADLELKEP--GEATGATNIVRALERPLKQIFENGLEGVAAEKVRNSPAGTGLNAA 477
 QY 481 TLSYGDLAADVDPVKYTRSAVLNASSVARNVLTTEVVDKPAKA 527
 DB 478 TGVYDMLAEGILDPKAYTRSAVLNASSVARNVLTTEVVDKPEKA 524

RESULT 8

ID CH62_MYCLE STANDARD: PRT: 540 AA.

AC P09239;
 DT 01-MAR-1989 (Rel. 10, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE 60 KDA CHAPERONIN 2 (PROTEIN CPN60 2) (GROEL PROTEIN 2) (65 KDA
 DE ANTIGEN).
 GN GROEL-2 OR MCBI1450.05C.
 OS Mycobacterium leprae.
 CC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 CC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.

CC SEQUENCE FROM N.A.
 CC MEDLINE=96313701; PubMed=2428046;
 CC Mehra V., Sweetser D., Young R.A.;
 CC "Efficient mapping of protein antigenic determinants.";
 CC Proc. Natl. Acad. Sci. U.S.A. 83:7013-7017(1986).
 CC [12]
 CC RP SEQUENCE FROM N.A.
 CC Brown D., Churcher C.M., James K.D., Parkhill J., Barrell B.G.,
 CC Rajandream N.A. (1998) to the EMBL/GenBank/DBJ databases.
 CC Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: PREVENTS MISFOLDING AND PROMOTES THE REFOLDING AND
 CC PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED UNDER STRESS
 CC CONDITIONS (BY SIMILARITY).
 CC -1- SUBUNIT: OLIGOMER OF 14 SUBUNITS COMPOSED OF TWO STACKED RINGS OF
 CC 7 SUBUNITS (BY SIMILARITY).
 CC -1- DISEASE: PURIFIED 65 KDA ANTIGEN CAN ELICIT A STRONG DELAYED-TYPE
 CC HYPERSENSITIVITY REACTION IN EXPERIMENTAL ANIMALS INFECTED WITH
 CC MYCOBACTERIUM TUBERCULOSIS.
 CC -1- MISCELLANEOUS: THIS PROTEIN IS ONE OF THE MAJOR IMMUNOREACTIVE
 CC PROTEINS OF THE MYCOBACTERIA. THIS ANTIGEN CONTAINS EPITOPES THAT
 CC ARE COMMON TO VARIOUS SPECIES OF MYCOBACTERIA.
 CC -1- SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.


```

Db      301 DLGELKOTTFVQIGRARVYKQKENTIIIVDASDPKEIQKRIASIKQSIIEETITDPEDE 360
Qy      361 KUGERLAKTAGVAVYIKYGAATDTALKEKESVEDAVAAAKAABEGVGGASLIHOA 420
Db      361 KUGERLAKTAGVAVYIQVGAATETEMKEKKLIEDLATAKFAVEGCIYAGGGTALVNYI 420
Qy      421 RAALTELARLQDELGVDFSGVDSLAAPLAFMIANAGIDGSSVYVKNKSELPAGGLNAN 480
Db      421 PK-VAALDTFVSGDEDTGQOITLRALPEEPVROIAENAGLESSIVEKKAASEPGIGFDAY 479
Qy      481 TSYGDLADAGVIDPYKVRSAVLNASSVARYAVLTETETVYVAPK 526
Db      480 NEKYVMIEAGIVDPKAVYTRSLQNAASVASMVLTESVVAIDPEK 525

RESULT 10
CH60_THEBR STANDARD: PRT; 540 AA.
NC 060024:
NC 01-NOV-1997 (Rel. 35, Created)
NC 01-NOV-1997 (Rel. 35, Last sequence update)
NC 30-MAY-2000 (Rel. 39, Last annotation update)
Df 60 KDA CHAPERONIN (PROTEIN CPN60) (GROEL PROTEIN).
Dt MOPA OR GROEL.
Os Thermoaerobacter brockii (Thermoaerobium brockii).
Cc Bacteria; Firmicutes; Bacillus/Clostridium group;
Cc Thermoaerobacter group; Thermoaerobacter.
RN 11)
RN SEQUENCE FROM N.A.
RC STRAIN=RT8.G4;
RC MEDLINE=99014232; PubMed=9795109;
RA Truscott K.N., Scopes R.K.:
RT "Sequence analysis and heterologous expression of the groE genes from
RT Thermoaerobacter sp. Rt8.G4."
RT Gene 217:15-23(1998).
RN 12)
RN SEQUENCE OF 1-43, AND CHARACTERIZATION.
RP STRAIN=RT8.G4;
RC MEDLINE=94291621; PubMed=7912671;
RA Truscott K.N., Hoef P.B., Scopes R.K.:
RT "Purification and characterization of chaperonin 60 and chaperonin 10
RT from the anaerobic thermophile Thermoaerobacter brockii."
RL Eur. J. Biochem. 222:277-284(1994).
CC -1- FUNCTION: PREVENTS MISFOLDING AND PROMOTES THE REFOLDING AND
CC PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED UNDER STRESS
CC CONDITIONS (BY SIMILARITY).
CC -1- SUBUNIT: OLIGOMER OF 14 SUBUNITS COMPOSED OF TWO STACKED RINGS OF
CC 7 SUBUNITS (BY SIMILARITY).
CC -1- MASS SPECTROMETRY: MW=5794.9; MW ERR=10; METHOD=ELECTROSPRAY.
CC -1- SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-slb.ch/announce/
CC or send an email to license@isb-slb.ch).
CC -----
CC DR EMBL, U56021; IAB00559.1; -.
CC DR HSSP, P06139; IGR1.
CC DR INTERPRO: IPR001844; -.
CC DR INTERPRO: IPR002423; -.
CC DR PFAM: PF00118; cpn60_TCP1.1.
CC DR PRINTS: PR00288; CHAPERONIN60.
CC DR PRINTS: PR00304; TCOMPLEXTCP1.
CC DR PROSITE: PS00296; CHAPERONINS_CPN60; 1.
CC KW Chaperone; ATP-binding.
CC FT INIT MET 0
CC FT 43 43 G -> A (IN REF. 2).
CC FT CONFLICT 43 43
CC SEQUENCE 540 AA: 57928 MW; 27C316AD35C61734 CRC64;

```


RA Taguchi H., Konishi J., Ishii N., Yoshida M.;
 RT "A chaparrin from a thermophilic bacterium, *Thermus thermophilus*,
 RT that controls refolding of several thermophilic enzymes.";
 RL J. Biol. Chem. 266:22411-22418(1991).
 RN [4]
 RP X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS) OF 190-334.
 RA Walsh M.A., Dementieva I., Evans G., Sanishvili R., Joachimiak A.;
 RL Submitted (MAR-1999) to the PDB data bank
 CC -1- FUNCTION: PREVENTS MISFOLDING AND PROMOTES THE REFOLDING AND
 CC PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED UNDER STRESS
 CC CONDITIONS.
 CC -1- SUBUNIT: OLIGOMER OF 14 SUBUNITS COMPOSED OF TWO STACKED RINGS OF
 CC 7 SUBUNITS (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; DA5880; BAA08299.1; -
 CC EMBL; U29483; AAA83441.1; -
 CC PIR; A39313; A39313.
 CC PDB; 1SRV; 12-MAR-99.
 CC INTERPRO; IPR001844; -
 CC INTERPRO; IPR002423; -
 CC PFM; PF00118; cpn60_TCP1; 1.
 CC PRINTS; PR00298; CHAPERONIN60.
 CC PRINTS; PR00304; TCOMPLEXTCP1.
 CC PROSITE; PS00296; CHAPERONINS_CPN60; 1.
 CC Chapterone; ATP-binding; 3D-structure.
 CC INIT_MET 0
 CC CONFLICT 34 34 R -> N (IN REF. 3).
 CC CONFLICT 238 238 R -> R (IN REF. 2).
 CC CONFLICT 436 439 TGAK -> RAPE (IN REF. 2).
 CC SEQUENCE 542 AA; 57758 MW; 1A3C991A893DE8E CRC64;
 SO
 Query Match 53.7%; Score 1426.5; DB 1; Length 542;
 Best Local Similarity 54.4%; Pred. No. 1.4e-61;
 Matches 289; Conservative 100; Mismatches 137; Indels 5; Gaps 3;

QY 421 RKAETELRSLTGDDEVLDVDFSEALAPLFMTAANAGIDGSVYVVKV---SFLPAGHGL 477
 DB 420 ISAVEELIKLEBDEATGAKIVRALEEPARQIAENAGEGSVYVQOILAEKRNPR-YGF 478
 QY 478 NVNTLSYGLAADGVIDPVKTSRAVLAASSVARNVLTETVVDPAKAE 528
 DB 479 NMTGGEVDMVEAGIYDPAKVTTSALQNASIGALILITDAVVAEKPKE 529
 RESULT 12
 CH60_BACST STANDARD; PRT; 539 AA.
 ID CH60_BACST
 AC 007201;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE 60 KDA CHAPERONIN (PROTEIN CPN60) (GROEL PROTEIN).
 GN MOPA OR GROEL.
 OS Bacillus stearothermophilus.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NUB36;
 RX MEDLINE=93224474; PubMed=8096841;
 RA Schoen U., Schumann W.;
 RT "Molecular cloning, sequencing, and transcriptional analysis of the
 RT groEL operon from *Bacillus stearothermophilus*.";
 RL J. Bacteriol. 175:2465-2469(1993).
 CC -1- FUNCTION: PREVENTS MISFOLDING AND PROMOTES THE REFOLDING AND
 CC PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED UNDER STRESS
 CC CONDITIONS (BY SIMILARITY).
 CC -1- SUBUNIT: OLIGOMER OF 14 SUBUNITS COMPOSED OF TWO STACKED RINGS OF
 CC 7 SUBUNITS (BY SIMILARITY).
 CC -1- INDUCTION: BY HEAT SHOCK.
 CC -1- SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; L10132; AAA22752.1; -
 CC DR HSSP; P06139; 1GRL.
 CC DR INTERPRO; IPR001844; -
 CC DR INTERPRO; IPR002423; -
 CC PFM; PF00118; cpn60_TCP1; 1.
 CC PRINTS; PR00298; CHAPERONIN60.
 CC PRINTS; PR00304; TCOMPLEXTCP1.
 CC PROSITE; PS00296; CHAPERONINS_CPN60; 1.
 CC Chapterone; ATP-binding; Heat shock.
 CC SEQUENCE 539 AA; 57244 MW; 37FE4ADD7492CC3E CRC64;
 SO
 Query Match 53.3%; Score 1415.5; DB 1; Length 539;
 Best Local Similarity 54.0%; Pred. No. 4.6e-61;
 Matches 283; Conservative 98; Mismatches 140; Indels 3; Gaps 1;


```

CC -----
DR EMBL: X62914; CAA44697.1; -
DR PIR: S18869; S18869.
DR PIR: S22342; S22342.
DR HSP: P06139; 1GRU.
DR INTERPRO: IPR001844; -.
DR INTERPRO: IPR002423; -.
DR PFAM: PF00118; cpn60_TcPI.1.
DR PRINTS: PR00298; CHAPERONIN60.
DR PRINTS: PR00304; TCOMPLEXTCP1.
DR PROSITE: PS00296; CHAPERONIN60.1.
DR Chaperone; ATP-binding.
KW Chaperone; 539 AA; 57380 MW; D13AA4D40C8EC580 CRC64;
SQ
Query Match 52.7%; Score 1399; DB 1; Length 539;
Best Local Similarity 52.4%; Pred. No. 2.8e-60;
Matches 277; Conservative 109; Mismatches 135; Indels 8; Gaps 2;
Y 1 SKLEEDETARRAMEVGMOKLADPTVRYTLGPRRHVYLAAFGGPTVNGCVTARETEL 60
D 2 AKTLLEGEERKRSQAGVNDKLANVKTLCGPKGNVLDKFFSPPLTNDGVTIARETEL 61
QY 61 EDFEEDLGAOLVKSVAETKNDVAGDGTATTLAQAALIKGLRLVAGVPVALGVIGK 120
D 62 EDAYENMGALVKEVATKTDVAGDGTATTLAQAALIRGLKNTVAGAMPILIRNGIKT 121
QY 121 ADAVFALLASATPVSGKTGIAQVATVSSRDEQIGDLVGEAMNKVGHDSVSKESSTL 180
D 122 AVKAAVEEIQKISPKVKGEDIRVAALISADEKIGKILADAMEKVGNEGVITVEESKM 181
QY 181 GTELEFEIGIFPKGFSAVEFVTDQNAQVLEDPALTLHQDKISSLPDLPLEKAVGT 240
D 182 GTELDVEEGQFDGVSAMVDTMEKNAVLDMPVLITDKKISNODPLLEQVQA 241
QY 241 GKPLLYAEDVEGEALATLVNAIRKTLKAAVAGPYFGDRRAFLDELAIVTGHVNP 300
D 242 GKRLITLADIECEAMTTLVNNKRGFTCVGAKAPGFGRRKEMLDIATLGGVYISD 301
QY 301 DAGIVREVELEVIGSRRRVVSKDPTVIYDGGSTAFAVANRANHAEIDKSDSDMDRE 360
D 302 EVGGDLKEATLDMIGESVYKREESTIVNGRNSSEIRNINQIKLOEATTSERDKE 361
QY 361 KLGERTLAKGAVAVIKVGAATDTALKERKESYEDAAVAAKAAVEEGVGGG---ASLI 417
D 362 KIGERLAKLAGVAVVAVGATETELKESKRLIEDLAATKAAVEEIVGGGTAAYNVI 421
QY 418 HQARKALTELRLASTGDEVIGVDFSEBALAAPFWLTAANAGLDGSVVNVSELPAGHGL 477
D 422 NEVAKLTSDIQ-----DEQGINIVRSLEPRKQIAHNNAGLGSVILEKVKNSDAGVGF 476
QY 478 NVNTLTSLGLADAGVIDPVKVTNSAVINASSVAMVLTTEVVVVDKPK 526
D 477 DALRGEKDKIMKAGIVDPTKVTRSALONASVASTFELTEAAVADLPEK 525
DB
RESULT 15
CH60_BACSU STANDARD; PRT; 543 AA.
ID CH60_BACSU STANDARD; PRT; 543 AA.
AC P28598; 005526;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE 60 KDA CHAPERONIN (PROTEIN CPN60) (GROEL PROTEIN) (STRESS PROTEIN H5).
GN MOPA OR GROEL.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92283754; PubMed=1350777;
RA Schmidt A., Schlesswohl M., Voelker U., Hecker M., Schumann W.;
RT "Cloning, sequencing, mapping, and transcriptional analysis of the

```

```

RT groESL operon from Bacillus subtilis.";
RL J. Bacteriol. 174:3993-3999(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=92283753; PubMed=1350776;
RA Li M., Wong S.L.;
RT "Cloning and characterization of the groESL operon from Bacillus
subtilis.";
RL J. Bacteriol. 174:3981-3992(1992).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168 / MARBURG;
RX MEDLINE=93129852; PubMed=1369494;
RA Tozawa Y., Yoshikawa H., Kawamura F., Itaya M., Takahashi H.;
RT "Isolation and characterization of the groES and groEL genes of
Bacillus subtilis Marburg.";
RL Biosci. Biotechnol. Biochem. 56:1995-2002(1992).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=168 / MARBURG;
RX MEDLINE=98116660; PubMed=9455482;
RA Kasahara Y., Nakai S., Ogasawara N., Yata K., Sadate Y.;
RT "Sequence analysis of the groESL-cotA region of the Bacillus subtilis
genome, containing the restriction/modification system genes.";
RL DNA Res. 4:335-339(1997).
RN [5]
RP SEQUENCE OF 1-18 FROM N.A.
RC STRAIN=168 / MARBURG;
RX MEDLINE=97346038; PubMed=9202461;
RA Sadate Y., Yata K., Fujita M., Sagai H., Itaya M., Kasahara Y.,
RA Ogasawara N.;
RT "Nucleotide sequence and analysis of the phoB-rne-groESL region of
the Bacillus subtilis chromosome.";
RL Microbiology 143:1861-1866(1997).
RN [6]
RP SEQUENCE OF 1-30.
RC STRAIN=1558;
RX MEDLINE=94282319; PubMed=8012595;
RA Voelker U., Engelmann S., Maul B., Rietdorf S., Voelker A.,
RA Schmid R., Mach H., Hecker M.;
RT "Analysis of the induction of general stress proteins of Bacillus
subtilis.";
RL Microbiology 140:741-752(1994).
RN [7]
RP SEQUENCE OF 1-30.
RC STRAIN=1558;
RX MEDLINE=93123969; PubMed=1362210;
RA Voelker U., Mach H., Schmid R., Hecker M.;
RT "Stress proteins and cross-protection by heat shock and salt stress
in Bacillus subtilis.";
RL J. Gen. Microbiol. 138:2125-2135(1992).
CC -1- FUNCTION: PREVENTS MISFOLDING AND PROMOTES THE REFOLDING AND
PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED UNDER STRESS
CONDITIONS (BY SIMILARITY).
CC -1- SUBUNIT: OLIGOMER OF 14 SUBUNITS COMPOSED OF TWO STACKED RINGS OF
7 SUBUNITS (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).
CC EMBL: M84965; AAA22531.1; -
DR EMBL: M81132; AAA22503.1; -
DR EMBL: D10972; BAA22519.1; -
DR EMBL: AB007637; BAA22747.1; -
DR EMBL: D88802; BAA19727.1; -
DR EMBL: Z59107; CAB12422.1; -

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 5, 2001, 14:24:19 ; Search time 67.42 seconds
(without alignments)
541.836 Million cell updates/sec

Title: US-09-461-774-2

Perfect score: 2654
Sequence: 1 SKLIEYDETRARRAMEVGMK.....VYVDKPAKEDHDHGHGHAH 538

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 195891 seqs, 67900655 residues

otal number of hits satisfying chosen parameters: 195891

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : PIR_66:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB ID	Description
1	2602	98.0	539	2 F70737	chaperonin groEL1
2	2223	83.8	537	2 S25181	heat shock protein
3	1643.5	61.9	540	2 B41325	heat shock protein
4	1642	61.9	541	2 T35591	chaperonin cpn60 -
5	1620.5	61.1	540	2 S37566	groEL1 protein - S
6	1574.5	59.3	540	2 A26950	groEL2 protein - M
7	1574.5	59.3	540	2 A43509	65K antigen mbaa -
8	1543.5	58.2	540	2 C41325	heat shock protein
9	1543.5	58.2	541	2 S40245	heat shock protein
10	1543.5	58.2	541	2 T44725	chaperonin 60K (lm
11	1543.5	58.2	588	2 A25902	65K antigen - Myco
12	1454.5	54.8	541	2 S68249	chaperonin groEL h
13	1440	54.3	541	2 S72614	chaperonin 60 - Th
14	1423.5	53.6	544	2 JC6063	chaperonin groEL -
15	1422	53.6	538	2 H73677	groEL protein - Th
16	1410.5	53.1	539	2 B49855	heat shock protein
17	1402.5	52.8	548	2 G75499	groEL protein - De
18	1401	52.8	543	2 B41872	heat shock protein
19	1389	52.7	539	2 S22342	chaperonin HSP60 -
20	1398.5	52.7	544	2 B41884	chaperonin HSP60 -
21	1392.5	52.3	538	2 J04195	58K heat shock pro
22	1387	52.3	541	2 B44425	chaperonin groEL -
23	1385	52.2	546	2 S34938	chaperonin groEL -
24	1384.5	52.2	544	2 JC5130	heat shock protein
25	1369	51.6	552	2 S74322	chaperonin groEL-2
26	1365.5	51.5	543	2 S70013	chaperonin-like pr
27	1359	51.2	541	2 JN0512	heat shock protein
28	1344	50.6	544	1 BVCGL	chaperonin groEL -
29	1337	50.4	552	2 S39765	chaperonin 60 - Co

30	1329.5	50.1	542	2 JN061	heat shock protein
31	1329.5	50.1	542	2 S32106	groEL protein - La
32	1328.5	50.1	545	2 JN0509	heat shock protein
33	1322.5	49.8	546	2 S65596	heat shock protein
34	1321.5	49.8	546	2 I40342	heat shock protein
35	1321.5	49.8	547	2 I40331	cpn60 protein (Gro
36	1320.5	49.8	546	2 S22347	groEL - Brucella a
37	1318.5	49.7	544	2 H81964	chaperonin 60K su
38	1318.5	49.7	546	2 B47073	chaperonin groEL -
39	1318.5	49.7	547	2 B83098	GroEL protein PA3
40	1318	49.7	539	2 JN0601	heat shock protein
41	1317.5	49.6	544	2 C81021	chaperonin, 60 kDa
42	1316.5	49.6	545	2 C70489	GroEL - Aquifex ae
43	1315.5	49.6	544	2 S23918	groEL protein - Ag
44	1314.5	49.5	544	2 S37039	groEL protein - Ba
45	1311.5	49.4	544	2 B36917	heat shock protein

ALIGNMENTS

RESULT 1

F70737 chaperonin groEL1 - Mycobacterium tuberculosis (strain H37RV)

N:Alternate names: 65K antigen; chaperonin-60; KCS protein

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000

C:Accession: F70737; B47292; S18041

R:Colo, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon

Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno

A:Reference number: A70500; MUID:98295987

A:Accession: F70737

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-539 <COL>

A:Cross-references: GB:277165; GB:AL123456; NID:g3261609; PIDN:CA01006.1; PID:g14493

A:Experimental source: strain H37RV

R:Kong, T.H.; Coates, A.R.; Butcher, P.D.; Hickman, C.J.; Shinnick, T.M.

Proc. Natl. Acad. Sci. U.S.A. 90, 2608-2612, 1993

A:Title: Mycobacterium tuberculosis expresses two chaperonin-60 homologs.

A:Reference number: A47292; MUID:93219332

A:Accession: B47292

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-466, 'K', 468-539 <KON>

A:Cross-references: EMBL:X60350; NID:g44599; PIDN:CA042909.1; PID:g44601

A:Note: sequence extracted from NCBI backbone (NCBIN:128605, NCBI:128608)

C:Genetics:

A:Gene: groEL1

C:Superfamily: chaperonin groEL

C:Keywords: molecular chaperone

Query Match 98.0%; Score 2602; DB 2; Length 539;

Best Local Similarity 98.1%; Pred. No. 2e-134;

Matches 528; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY	1	SKLIEYDETRARRAMEVGMKLDIVTAVRTIGPRRRHYVLAKAFGCPVTNDGVVAAREIEL	60
DB	2	SKLIEYDETRARRAMEVGMKLDIVTAVRTIGPRRRHYVLAKAFGCPVTNDGVVAAREIEL	61
QY	61	EDPFEDLGAQLVSVATKTNDVAGDGTATTAQALINGGLVLAAGVNPVALGVDICK	120
DB	62	EDPFEDLGAQLVSVATKTNDVAGDGTATTAQALINGGLVLAAGVNPVALGVDICK	121
QY	121	AADAVEALLASATPVSGKTGIAQAVTVSSRDQIDGLVGEAMNKVGHDSVSKESSTL	180
DB	122	AADAVEALLASATPVSGKTGIAQAVTVSSRDQIDGLVGEAMNKVGHDSVSKESSTL	181

QY	181	GELEFTEFGIGKHKEFLSAFYETDEDNQAVLEDLAILLHOQISSLPPLLEKRVAGT	240
Db	182	GTELEFTEFGIGDKKFLSAFYETDEDNQAVLEDLAILLHOQISSLPPLLEKRVAGT	241
QY	241	GKPLLIIVADEGEGELATLVNVAIKRTKLAAVAKCPYEGDRKAFLEDLAVVTTGGHHVNP	300
Db	242	GKPLLIIVADEGEGELATLVNVAIKRTKLAAVAKCPYEGDRKAFLEDLAVVTTGGQVNP	301
QY	301	DAGIVLREAGLEVLASARVVVSKDDTVYIVDGGTAENAVANANFLRAETDKSDSDMDPE	360
Db	302	DAGWVLREAGLEVLASARVVVSKDDTVYIVDGGTAENAVANANKILRAETDKSDSDMDRE	361
QY	361	KLGEFLAKIAGVANIYKGAATDTLALKEKESVEDAIVAAAKAAVEEGIYPPGGASLTHOA	420
Db	362	KLGEFLAKIAGVANIYKGAATETALKEKESVEDAIVAAAKAAVEEGIYPPGGASLTHOA	421
QY	421	RKALTELRASLTGDGDEVLGADVFSSEALAAPLFIWIAANAGIDGSSVVNNKYSSELPAGHGLVNN	480
Db	422	RKALTELRASLTGDGDEVLGADVFSSEALAAPLFIWIAANAGIDGSSVVNNKYSSELPAGHGLVNN	481
QY	481	TLTSGDILAADGYITDPYKKTARSANVLAASSVARKVLTTEYIVVNDKPKAKEDHNNHNGHAN	538
Db	482	TLTSGDILAADGYITDPYKKTARSANVLAASSVARKVLTTEYIVVNDKPKAKEDHNNHNGHAN	539

```

RESULT 2
S25181
heat shock protein groEL - Mycobacterium leprae
N:Alternate names: protein B29_C3_248
C:Species: Mycobacterium leprae
C:Date: 03-Feb-1994 #sequence_revision 03-Feb-1994 #text_change 21-Jul-2000
C:Accession: S25181, S72819, S72998
R:Id wit. T.F.R.; Bekele, S.; Oslund, A.; Miko, T.L.; Hermans, P.W.M.; van Soelingen, I
Mol. Microbiol. 6, 1995-2007, 1992
A:Title: Mycobacteria contain two groEL genes: the second Mycobacterium leprae groEL gene
A:Reference number: S25180; MUID:92374850
A:Accession: S25181
A:Molecule type: DNA
A:Residues: 1-537 <DEUX>
A:Cross-references: EMBL:Z11665; NID:q44408; PIDN:CAB62574.1; PID:q6562688
A:Genetics: GI

```

A:Description: Mycobacterium lepreae cosmid B1620.
A:Reference number: S72580
A:Accession: S72819
A:Status: preliminary
Molecule type: DNA
Residues: 1-500 <SMI>
A:Cross-references: EMBL:U00015
A:Experimental source: cosmid B1620
A:Genetics: G2
R:Smith, D.R.; Robison, K.
submitted to the EMBL Data Library, November 1993
A:Description: Mycobacterium lepreae cosmid B229.
A:Reference number: S72580
A:Accession: S72998
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-537 <SMW>
A:Cross-references: EMBL:U00020; NID:g467102; PIDN:AAA17312.1; PID:g467130
A:Experimental source: cosmid B229
A:Genetics: G2
C:Genetics: <GI>
A:Gene: groEL
A:Note: this groEL gene is arranged in an operon with groES
C:Genetics: <G2>
A:Gene: groEL
A:Note: cosmids B1620 and B229
C:Superfamily: chaparonin groEL
Keywords: molecular chaparonine

Query Match	83.88;	Score 2223;	DB 2;	Length 537;
Best Local Similarity	82.38;	Pred. No. 7.9e-114;		
Matches 443; Conservative	48;	Mismatches 45;	Indels ?;	Gaps 2;

QY 1 SKLIEDELFARRAMEGMDKLADTVAVTLGPGGRHNVLLAKAGSPVTVNDGVTVAHEEL 60
DB 2 SKLIEDELFARRAMEGMDKLADTVAVTLGPGGRHNVLLAKAGSPVTVNDGVTVAHEEL 61
QY 61 EDPEFDLGAQLVSVATKTNVDAGDGTATTLAQLIKGLRLVAAGVPAALGVGIC 120
DB 62 EDPEFDLGAQLVSVATKTNVDAGDGTATTLAQLIKGLRLVAAGVPAALGVGIC 121
QY 121 AADVAFEALLASATPVSGKTGTAAVATVSSRPDQIGDIVGEAMNNKVGHSVSVKESSTL 180
DB 122 AADVAFEALLASATPVSGKTGTAAVATVSSRPDQIGDIVGEAMNNKVGHSVSVKESSTL 181
QY 181 GTELEFTEGIGPHKGLSLAFVTFDEPMQAVLEDALELILHODKISSLPDLPLELEKACT 240
DB 182 DPELEFTEGIGPHKGLSLAFVTFDEPMQAVLEDALELILHODKISSLPDLPLELEKACT 241
QY 241 GPRLLIVADVGEALATLVVAIKRKLKAAVAKGPFEDRRKAPLEDAVUTGCHVNP 300
DB 242 GPRLLIVADVGEALATLVVAIKRKLKAAVAKGPFEDRRKAPLEDAVUTGCHVNP 301
QY 301 DAGIVLREGLLEVLSARVVVSKDDTVLVDSGGTAEAVANRANHLRAEIDKSDSDMRE 360
DB 302 EGVGLVREGLLEVLSARVVVSKDDTVLVDSGGTAEAVANRANHLRAEIDKSDSDMRE 361
QY 361 KLGERLAKIAGGAVAIKVGAAATDTALKEKESVEDAVAAAKAAVEEGIVGGASLIHQ 420
DB 362 KLGERLAKIAGGAVAIKVGAAATDTALKEKESVEDAVAAAKAAVEEGIVGGASLIHQ 420
QY 421 KRALTELRLSLTGDDEVLCGVDFSELAALPLMTAANAGIDGSVVVNVKVELPAGCLANN 480
DB 421 KRALTELRLSLTGDDEVLCGVDFSELAALPLMTAANAGIDGSVVVNVKVELPAGCLANN 480
QY 481 TTSYGDLDADGVITDEPKVTRSAVLAASSVAAVAVLTFTETVYVVKPAKAEEDHNNHGH 538
DB 481 TTSYGDLDADGVITDEPKVTRSAVLAASSVAAVAVLTFTETVYVVKPAKAEEDHNNHGH 537
RESULT 3
B41325
heat shock protein 58 - Streptomyces albus
N:Alternate names: heat shock protein groEL homolog 1
C:Species: Streptomyces albus
C:Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 26-Aug-1999
C:Accession: B41325
R:Mazodler, P.; Guglielmi, G.; Davies, J.; Thompson, C.J.
J. Bacteriol. 173, 7382-7386, 1991
A:Title: Characterization of the groEL-like genes in Streptomyces albus.
A:Reference number: A41325; MUID:92041639
A:Accession: B41325
A:Molecule type: DNA
A:Residues: 1-540 <MAZ>
A:Cross-references: GB:M76657; NID:g153290; PID:AAA26753.1; PID:g153292
C:Genetics:
A:gene: groEL1
C:Superfamily: chaperonin groEL

	Query Match	61.9%	Score 1643.5;	DB 2,	Length 540;
	Best Local Similarity	61.0%;	Pred. N.2.6e-82;		
	Matches 330;	Conservative 87;	Mismatches 119;	Indels 5;	Gaps 4;
QY	1 SKLIYEDERARAMEVGMKDLADYRYVLGPRGRHVLAKACGGPTVNNDGTYVARRETEL	60			
	: : : : : : : : : : : : : : : : : : : :				
Db	2 AKILFDEBARALREGVNLADYTKVTIGPGRNVNIDKKFGAPITINDGTIAREVC	61			
QY	61 DDPFDICAGQIVSVATKTNDVAGSGTTATLLAOLIKGSLRVAAGVNPALGVGIGK	120			
	: : : : : : : : : : : : : : : : : : : :				
Db	62 DDPIYNLGAQLVKEVATTKTNDAGSGTTATVLAALVREGSRNNVAAAGSPALKKGIDA	121			

[illegible]

OY	239	GTRKLLVADVEBESALATLVNAIKRTLKAVAVKGPFDORRKAPLEDLAVVTGGHV	298
		: : : : :	
Db	242	NASKELLIAEDLEBEALSTLVNKRIGTFANVAWVKAPGCFDRKKAMJOMAVLTGTVI	301
		: : : : :	
OY	299	NPDACIVIREVGLEYLSARRVVSKDQTVVDGGTAEAANPANNILRAEIDKSQSDMD	358
		: : : : : : : : :	
Db	302	SEEWELKIDOVGLEVLGTARRITVTKDDTIVDSAGRDEVOGRIOAIKAEIENTDSDMD	361
		: : : : : : : : :	
OY	359	REKLIERLAKLAGAAVAVTKGAADPTALKERKESEDVAANAARAAKEEGIVPGGASLIH	418
		: : : : :	
Db	362	REKLOERLAKLAGVCYCVTKVAATEVELKERKHLEDIASITRAAVEGIVSGGSALVH	421
		: : : : :	
OY	419	QARKALTELRLASTGDEVLYGVDFESEALAPLFWIIANAGLDGSSVVNNKSELPRAGHLN	478
		: : : : :	
Db	422	AVK - VLEGNGLGKDEATGVAAYVRRAVEPLRIIAENAGLEGVITISXAADDKOGGN	479
		: : : : :	
OY	479	VNTLSTYGDLAADGYIDPVKVTBSAVLNASSYARNVLTTETVVYVDKPAAKD -- HDHHGH	536
		: : : : :	
Db	480	AATGEYGDVLKAGVIDPYKVTRSALENNASIASLTLTTLETVLVERKKEEPBPAAGSHSH	539
		: : : : :	
OY	537	AH 538	
Db	540	SH 541	

```

Db      480  AATGCGYDLVKAGVADPYKVTTRSALFNMAASIASLLTTFETLVEKRAEPEEAGHGHS 539
Oy      538  H 538
Db      540  H 540

RESULT  4
T35591
chaperonin cpn60 - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 11-Jan-2000
C:Accession: J35591
R:Saunders, D.; Harris, D.; Parthill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, August 1998
A:Reference number: J21583
A:Accession: J35591
A:Status: Preliminary; translated from GR/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-541 <SAU>
A:Cross-references: EMBL:AL031317; PIDN:CAA20418.1; GSPDB:GN00070; SCQEDB:SC664
A:Experimental source: strain A3(2)
A:Genetics:
C:Gene: groEL1, SCQEDB:SC664.40
C:Superfamily: chaperonin groEL

Query Match      61.9%; Score 1642; DB 2; Length 541;
Best Local Similarity 60.3%; Pred. No. 3,1e-82;
Matches 327; Conservative 92; Mismatches 117; Indels 6; Gaps 3;

Oy      1  SKLIYDEFARAMEVGMGDKLADTVRVTLGPRGRHVVLAKARGPPTVTTNDGVTARTEL 60
      1 ::::|||||:|:::|||||:|:::|||||:|  |||:|||||:||||:|
Db      2  AKILAFDEEDARRALERGVAKLADIVKVTIIGPKRNVVIDKKTGAFTITNDGVTAREVEV 61
      2 ::::|||||:|:::|||||:|:::|||||:|  |||:|||||:||||:|

Oy      61  EDPFDLGAQILVKSVAFTKNDVAGDGTATTILAQALLKGLRLVAAGVNPALGVIGK 120
      61 :||:|||||:|||||:|||||:|||||:|||||:|  |||:|||||:||||:|
Db      62  EDPYENLGAQLVEKATKTNIDAGDGTATTATVLAQALVREGLKNTVAAGAASPALLKKGIDA 121
      62 :||:|||||:|||||:|||||:|||||:|||||:|  |||:|||||:||||:|

Oy      121  AADAVFEALIASATVSGKGTGAQVAVYSSRBEQIGDLVGEAMKNVGHDSVSVKESSTL 180
      121 :||:|||||:|||||:|||||:|||||:|||||:|  |||:|||||:||||:|
Db      122  AVAAASEDLTAAFRARIDKESDIAAAVAASDAQOQVGGELLAEAMDVVGKDGVTIVESNTF 181
      122 :||:|||||:|||||:|||||:|||||:|||||:|  |||:|||||:||||:|

Oy      181  GTELEFTEGIGKGFSLAFTVDPDNOCAVLEDAIILHOKRISPLPLPLEKRV-A 238
      181 :||:|||||:|||||:|||||:|||||:|||||:|  |||:|||||:||||:|
Db      182  GLELDFTEGMADKGLSLSYFTYTDDERMEAVLDDPYIILINQKRISSIDLPLLEKRVIA 241
      182 :||:|||||:|||||:|||||:|||||:|||||:|  |||:|||||:||||:|

```

[illegible]

Db 361 REKQERLAKLAGVCVKGAAATEVELKERHLEDAISATRAAVEGIVSGGSALVH 420
Qy 419 QARKALTELRLASTGDEVLGVDFSEALAPLFWIAANAGLDGSVVNVKSELPAHGILN 478
Db 421 AVK--VLEGNGKGTGDEATGTAAYVRRAAVEPLRKIAENAGLEGIVITSKADLDKGGQFN 478
Qy 479 VNTLSYDGLAADGVDPVKVTRSAVLNASSVARMVLTETTVVVDKPAKAD--HDHHHG 536
Db 479 AATGEYDGLVAGVIDPVKVTRSALENNAASIASLTLTETLVEKEEPEAPAGCHSHGH 538
Qy 537 AH 538
Db 539 SH 540
RESULT 6
A26950
groEL2 protein - Mycobacterium tuberculosis (strain H37RV)
N:Alternate names: 65K antigen
C:Species: Mycobacterium tuberculosis
Date: 02-Jun-1988 #sequence_revision 02-Jun-1988 #text_change 20-Jun-2000
Accession: A26950; MUID:87137260
R:Stimnick, T.M.
J. Bacteriol. 169, 1080-1088, 1987
A:Title: The 65-kilodalton antigen of Mycobacterium tuberculosis.
A:Reference number: A26950; MUID:87137260
A:Accession: A26950
A:Molecule type: DNA
A:Residues: 1540 <SH1>
A:Cross-references: GB:M15467; NID:g149999; PIDN:AAA8232.1; PID:g150000
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holtroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A:Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987
A:Accession: A70830
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1540 <COL>
A:Cross-references: GB:AL021932; GB:AL134456; NID:g3261527; PIDN:CAA17397.1; PID:g290951
A:Experimental source: strain H37Rv
C:Genetics:
A:Gene: groEL2; RV0440
C:Superfamily: chaperonin groEL
Query Match 59.3%; Score 1574.5; DB 2; Length 540;
Best Local Similarity 61.2%; Pred. No. 1.5e-78;
Matches 322; Conservative 75; Mismatches 126; Indels 3; Gaps 2;
Qy 1 SKLIEYDETRARAMEVGMKDLADTVRYTLGPRGHVYLAKAFGSPYTNDCVYVAREIEL 60
Db 2 AKTAYDEEARRGLEKRLNALDAVKYTLGPKGNVLEKRWGAPITTNCGVSTAKIEL 61
Qy 61 EDPEDGADLVKSVATKTDVAGDGTATTAALIKGSLRYAAGVNPVALGVGIGK 120
Db 62 EDPPEKIGAEVKEVAKTDDVAGDGTATTAALIKGSLRYAAGVNPVALGVGIGK 121
Qy 121 AADAVFALLASATPVSGKTGIAOVATVSSRDEQIGDLVGBAMNKVGHDSVYKESSTL 180
Db 122 AVEKVTETLLKGAKEVETKQIAATAISAGDOSIGDLIAEAMKVNEGIVTVEESNTF 181
Qy 181 GTELEFTEGIGFHKGFASAFVYDFDNOQAVLEBALILLHODKISSLPDLLPLEKYGAGT 240
Db 182 GLOLELTEGKRFKGYISGFVYDFDPERQEAILEDPYILLVSSKVSFKVDLPLEKYIGA 241
Qy 241 GKPLLIYAEEVEGALATLVVNAIRKTLKAVAVGYPFGRRKAFLEDLAVVVGCHVNP 300
Db 242 GKPLLIYAEEVEGALSTLVVNAIRKTFKSAVAKAPGRRKMLDMMALITGGQVISE 301
Qy 301 DAGIVLREVGLEVLGSARRVVS KDPTVIVDGGCTAAEVANRANHLRAETDKSDSDMRE 360

Db 302 EVGLTLENADLSLLGKARKVVTETKDETTIVEGAGDDTAIGRAVQIQLQEIENSDDVDRE 361
Qy 361 KLGRLAKLAGGVAVIVGATDTALKERKESVEDAAAKAAVEEIVGGGASLIHOA 420
Db 362 KLOERLAKLAGGVAVIKAGATEVELKERHRIEDAVRNKAAEEBIVAGGVTLL-QA 420
Qy 421 RKALTELRLASTGDEVLGVDFSEALAPLFWIAANAGLDGSVVNVKSELPAHGILNVN 480
Db 421 APTLDELK--LEGEATGAINIVKALAPLKQIAFNGLPEPGVAEVRNLPAGHGLNAQ 478
Qy 481 TLSYDGLAADGVDPVKVTRSAVLNASSVARMVLTETTVVVDKPAK 526
Db 479 TGYVEDLLAGVADPVKVTRSALQNAASIALGFLTEAVVADKPEK 524
RESULT 7
A43509
65K antigen mbaa - Mycobacterium bovis
C:Species: Mycobacterium bovis
Date: 28-Oct-1992 #sequence_revision 28-Oct-1992 #text_change 26-Aug-1999
Accession: A43509
R:Thole, J.E.R.; Keulen, W.J.; Kolk, A.H.J.; Groothuis, D.G.; Bervald, L.G.; Tiesjema
Infect. Immun. 55, 1466-1475, 1987
A:Title: Characterization, sequence determination, and immunogenicity of a 64-kilodalton
A:Reference number: A43509; MUID:87193155
A:Accession: A43509
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1540 <THO>
A:Cross-references: GB:M17705; NID:g149933; PIDN:AAA25358.1; PID:g149934
C:Superfamily: chaperonin groEL
Query Match 59.3%; Score 1574.5; DB 2; Length 540;
Best Local Similarity 61.2%; Pred. No. 1.5e-78;
Matches 322; Conservative 75; Mismatches 126; Indels 3; Gaps 2;
Qy 1 SKLIEYDETRARAMEVGMKDLADTVRYTLGPRGHVYLAKAFGSPYTNDCVYVAREIEL 60
Db 2 AKTAYDEEARRGLEKRLNALDAVKYTLGPKGNVLEKRWGAPITTNCGVSTAKIEL 61
Qy 61 EDPEDGADLVKSVATKTDVAGDGTATTAALIKGSLRYAAGVNPVALGVGIGK 120
Db 62 EDPPEKIGAEVKEVAKTDDVAGDGTATTAALIKGSLRYAAGVNPVALGVGIGK 121
Qy 121 AADAVFALLASATPVSGKTGIAOVATVSSRDEQIGDLVGBAMNKVGHDSVYKESSTL 180
Db 122 AVEKVTETLLKGAKEVETKQIAATAISAGDOSIGDLIAEAMKVNEGIVTVEESNTF 181
Qy 181 GTELEFTEGIGFHKGFASAFVYDFDNOQAVLEBALILLHODKISSLPDLLPLEKYGAGT 240
Db 182 GLOLELTEGKRFKGYISGFVYDFDPERQEAILEDPYILLVSSKVSFKVDLPLEKYIGA 241
Qy 241 GKPLLIYAEEVEGALATLVVNAIRKTLKAVAVGYPFGRRKAFLEDLAVVVGCHVNP 300
Db 242 GKPLLIYAEEVEGALSTLVVNAIRKTFKSAVAKAPGRRKMLDMMALITGGQVISE 301
Qy 301 DAGIVLREVGLEVLGSARRVVS KDPTVIVDGGCTAAEVANRANHLRAETDKSDSDMRE 360
Db 302 EVGLTLENADLSLLGKARKVVTETKDETTIVEGAGDDTAIGRAVQIQLQEIENSDDVDRE 361
Qy 361 KLGRLAKLAGGVAVIVGATDTALKERKESVEDAAAKAAVEEIVGGGASLIHOA 420
Db 362 KLOERLAKLAGGVAVIKAGATEVELKERHRIEDAVRNKAAVEEIVAGGVTLL-QA 420
Qy 421 RKALTELRLASTGDEVLGVDFSEALAPLFWIAANAGLDGSVVNVKSELPAHGILNVN 480
Db 421 APTLDELK--LEGEATGAINIVKALAPLKQIAFNGLPEPGVAEVRNLPAGHGLNAQ 478
Qy 481 TLSYDGLAADGVDPVKVTRSAVLNASSVARMVLTETTVVVDKPAK 526
Db 479 TGYVEDLLAGVADPVKVTRSALQNAASIALGFLTEAVVADKPEK 524


```
Db 242 GKRLVIAEDVEGEALATLLVNLKRGFTCVAAKAPGFGRRKAMLEDAIILTGCVITS 301
Qy 301 DAGIVREVLGVLSARRVYVSKDDTVIVDGGGTAEAVANRANHRAELDKSDMDRE 360
Db 302 DLELEKLDVTEVLEGRARQVOKEMNTIIVDAGDPEIKDIRIASIKSOIEETTSDFDRE 361
Qy 361 KLEBRILAKLAGVAIVKVGATDTALKERKESVEDAAVAAKAVEGIVGGASLIHQH 420
Db 362 KLEBRILAKLAGVAIVKVGATDETEMEKRLIEDLAATKAAVEGIVAGGITALVNI 421
Qy 421 RKALTELRASTLGVDFVSEALAPLFMTIAANAGLDGVSYYVVKSELPAHGILNVN 480
Db 422 PK-VAVAVLDTVSGDEKTVGIIIRALEEPYRQIAENANGLEGSVIVEKVASSEPGIGFDAY 480
Qy 481 TLSTYGLADGVIDPVKVTBSAVLNASSVARMVLTETTVVVDKPAK 526
Db 481 NEKYVMNIEAGIVDPKVTBSALONASVAVMLTTESVVDIPEK 526
```

RESULT 13

```
72614
Chapteronin 60 - Thermoanaerobacter brockii
N:Alternate names: heat shock protein groEL homolog
C:Species: Thermoanaerobacter brockii
C:Date: 14-Feb-1997 #sequence_revision 14-Feb-1997 #text_change 26-Aug-1999
C:Accession: S72614, S45615
R:Truscott, K.N.; Scopes, R.K.
Submitted to the EMBL Data Library, April 1996
A:Description: Cloning and sequence of the chapteronin operon of Thermoanaerobacter brockii
A:Reference number: S72613
A:Accession: S72614
A:Molecule type: DNA
A:Residues: 1-541 <TRW>
A:Cross-references: EMBL:U56021; NID:q1326189; PIDN:AA800599.1; PID:q1326191
R:Truscott, K.N.; Hoj, P.B.; Scopes, R.K.
Eur. J. Biochem. 222, 277-284, 1994
A:Title: Purification and characterization of chapteronin 60 and chapteronin 10 from the
A:Reference number: S45615; MID:94291621
A:Accession: S45615
A:Molecule type: protein
A:Residues: 2-34, 'X', 36-43, 'A' <TRW>
C:Genetics:
A:Gene: groEL
C:Superfamily: chapteronin groEL
C:Keywords: molecular chaperone
F:2-541/Product: chapteronin 60 #status experimental <MAT>
```

```
Query Match 54.3%; Score 1440; DB 2; Length 541;
Best Local Similarity 55.1%; Pred. No. 3e-71; Indels 8; Gaps 3;
Matches 292; Conservative 95; Mismatches 135;

Qy 1 SKLIEYDEFARRAMEVGMKLDATVRYTLGPRGRHVLAKAFGSPVTNDGVTVAREIEL 60
Db 2 AKGIKGEARALREGVANVADTVKTLGPRGRNVYLDKKYGSPLYTNGVITAREIEL 61
Qy 61 EDPFEDLGAOLVKSVAATKTDVAGDGTATTLAQAALIKGRLVAAGVNPVALGVGICK 120
Db 62 EDPFENAGQALLKEAATKTNDIAGDGTATTLAQAAMVREGKLNLAAGANPMILRGIRIAR 121
Qy 121 AAADAVEALLASATPVSGKGTIAQVATVSSRDEQIGDLVGEAMNKYGHDSVYKESSTL 180
Db 122 AADAVEALGRIRISKPIDNKESIAHVASISADEIKGLIAEAMDKVGKDGIVTEESKTL 181
Qy 181 GTLEFTEGIGFHKFGLSAFYVDFDNOQAVLEBDALILHODKISSLPDLPLEKVAQT 240
Db 182 GTLLEVEGNGQFDRGYASPTMYTDAEKMEAVLEPVLITDKKISNIGDLPLEQVVOQ 241
Qy 241 GRPLLVAEDVEGEALATLVNNAIRKTLKAVAAVGPFGDRRAFLIEDLAIVTGGHVVNP 300
Db 242 GKRLIIADVEGEALATLVNKLRTGFTCVAAKAPGFGRRKEMLODIAIILTGCVIASE 301
```

```
Qy 301 DAGIVREVLGVLSARRVYVSKDDTVIVDGGGTAEAVANRANHRAELDKSDMDRE 360
Db 302 ELGYDLKDVRLMDLGRARQVYKETTIVYGAGDSEIKRKNQKAOIEETTSYDRE 361
Qy 361 KLEBRILAKLAGVAIVKVGATDTALKERKESVEDAAVAAKAVEGIVPGGASLI 417
Db 362 KLEBRILAKLAGVAIVKVGATDETEMEKRLIEDLAATKAAVEGIVAGGITALVNI 421
Qy 418 HQARKALTELRASTLGVDFVSEALAPLFMTIAANAGLDGVSYYVVKSELPAHGILNVN 476
Db 422 EDVQKIVVD---SLEDGFKGAKIVRLALEEPYRQIAENANGLEGSVIVEKIKAKDPNFG 477
Qy 477 LNVNTLSYGLADGVIDPVKVTBSAVLNASSVARMVLTETTVVVDKPAK 526
Db 478 YDAVKEEFTDMKAGIVDPKVTBSALONASVAVMLTTESVVDIPEK 527
```

RESULT 14

```
JC6063
Chapteronin groEL - Bacillus sp.
C:Species: Bacillus sp.
C:Date: 31-Dec-1996 #sequence_revision 31-Dec-1996 #text_change 13-Sep-1998
C:Accession: JC6063
R:Xu, Y.; Zhou, P.J.
Acta Microbiol. Sin. 36, 241-249, 1996
A:Title: Phylogeny of molecular chaperone 60 proteins.
A:Reference number: JC6063
A:Accession: JC6063
A:Molecule type: DNA
A:Residues: 1-544 <XUA>
A:Experimental source: C-125 strain BD224
C:Genetics:
A:Gene: groEL
C:Superfamily: chapteronin groEL
C:Keywords: molecular chaperone
```

```
Query Match 53.6%; Score 1423.5; DB 2; Length 544;
Best Local Similarity 53.2%; Pred. No. 2.4e-70;
Matches 281; Conservative 109; Mismatches 135; Indels 3; Gaps 2;

Qy 1 SKLIEYDEFARRAMEVGMKLDATVRYTLGPRGRHVLAKAFGSPVTNDGVTVAREIEL 60
Db 2 AKDIKSEDAARRSMKLGVDKLDAAVAVTTLGPKRRNVYLEKKESSPLITNDGVITAREIEL 61
Qy 61 EDPFEDLGAOLVKSVAATKTDVAGDGTATTLAQAALIKGRLVAAGVNPVALGVGICK 120
Db 62 EDAPFNMGAKTLVAEVAASKTNDIAGDGTATTLAQAAMIRGLKNVTSGANPMVIRGRIGER 121
Qy 121 AAADAVEALLASATPVSGKGTIAQVATVSSRDEQIGDLVGEAMNKYGHDSVYKESSTL 180
Db 122 ATQVAVEELISKISKPIEGKKSIAQVATVSSADEVEKIIAEAMERKGNBGVITIEESKGP 181
Qy 181 GTLEFTEGIGFHKFGLSAFYVDFDNOQAVLEBDALILHODKISSLPDLPLEKVAQT 240
Db 182 STLEVEVEGNGQFDRGYASPTMYTDSKMEAVLEDPVYLITDKKISNIGDLPLEQVVOQ 241
Qy 241 GRPLLVAEDVEGEALATLVNNAIRKTLKAVAAVGPFGDRRAFLIEDLAIVTGGHVVNP 300
Db 242 GKPLIIAEDVEGEALATLVNKLRTGFTFNAVAVKAPGFGRRKAMLEDAIILTGGEVITE 301
Qy 301 DAGIVREVLGVLSARRVYVSKDDTVIVDGGGTAEAVANRANHRAELDKSDMDRE 360
Db 302 DGLDLKSANITGLGRSKVYVVKETTVIIEGAGESDKIAARNQKAOIEETTSDFDRE 361
Qy 361 KLEBRILAKLAGVAIVKVGATDTALKERKESVEDAAVAAKAVEGIVPGGASLIHQH 420
Db 362 KLEBRILAKLAGVAIVKVGATDETEMEKRLIEDLAATKAAVEGIVAGGITALVNI-V 420
Qy 421 RKALTELRASTLGVDFVSEALAPLFMTIAANAGLDGVSYYVVKSELPAHGILNVN 480
Db 421 IKAIVSIGAE--GDEATGVNIVRLALEEPYRQIAENANGLEGSVIVRLKKEAGFGFNAA 478
```


Db 2 AKTIAYDEARGLERGLNALADAVKVTLPGRNVLEKKMGAPTITNDGVSIAKEIETL 61
Qy 61 EDPFEDLGAOLVKSVAATNDVAGDGTATITLAOLIKGRLVAAGVNPALGVGIGK 120
Db 62 EDPYKICAEVLKVAKTDDVAGDGTATITLAOLVREGIRNVAAGANPLGLKRGIEK 121
Qy 121 AADAVEALLASAPFVSKGTGIAQVATVSSRDEQIGDLVGEAMKNGHDSVVSKESSL 180
Db 122 AVEAVTOSLKSAREVERKEQISATTAISAGDTQIGELIAEAMDKVNGEVIYESMTF 181
Qy 181 GTELEFTGIGFHKGFSLAYEPTDNOQAVLEDAIILLHODKISSLDLPLEKVAAGT 240
Db 182 GLOLELTGMRDKGYSIGYFTDAEROEALEDPYILLVSSKSTVADLPLEKVIQA 241
Qy 241 GKPLIIVAEDEVEGELATLVNAIRKTLKAVAVKPGYGRDKAPLEDAVVTGSHVNP 300
Db 242 GKPLIIVAEDEVEGELSTLVNKRIGTFKSAVAKPFGDRKAMLOMALLTGQVYSE 301
Qy 301 DAGIVLEVEGLVGSARRVVSCKDVTIVDGGTAAEVANRANHLRAEIKSDSDMDRE 360
Db 302 RVGLSLETAADVSLIGQARKVVTKDETTIVEGSDSDAIAGVQAIRAEIENSDDYDRE 361
Qy 361 KLGERLAKLAGVAVIKGAATDTRALKERKESVEDAVAAKAAVEEGIVPGGASLIHQ 420
Db 362 KLGERLAKLAGVAVIKGAATEVELKERKHRIEDAVARNAAVEEGIVAGGVALL-QS 420
Qy 421 RKALTELASLTGDEVLDVDFSEALAPLPFIANAGLDGSSVYVKNKSELPAGHGLVN 480
Db 421 APALDDL--GLTGDEATGANIVRALSLAPLQIAFNGLEGEVVAEKSNLPAGHGLNAA 478
Qy 481 TLSYGDLAADVDPVKYTRSAVLNASSVARNVLTETVVDKPAKA 527
Db 479 TGEYEDLKAGVADPVKTRSAIQNASIALFLTTEAVVADKPEKA 525

RESULT 2
US-08-997-362-160
Sequence 160, Application US/08997362
Patent No. 5985287
GENERAL INFORMATION:
APPLICANT: Tan, Paul
APPLICANT: Hiyma, Jun
APPLICANT: Visser, Elizabeth
APPLICANT: Skinner, Margot
APPLICANT: Scott, Linda
APPLICANT: Prestidge, Ross
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR
TREATMENT AND DIAGNOSIS OF MYCOBACTERIAL INFECTIONS
NUMBER OF SEQUENCES: 194
CORRESPONDENCE ADDRESS:
ADDRESSEE: Law Offices of Ann W. Speckman
STREET: 2601 Elliott Avenue, Suite 4185
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98121
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/997,362
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: U.S. Patent Application No. 5985287 08/973,970
FILING DATE: June 12, 1997
APPLICATION NUMBER: U.S. Patent Application No. 5985287 08/705,347
FILING DATE: August 29, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Sleath, Janet
REGISTRATION NUMBER: 37,007

REFERENCE/DOCKET NUMBER: 11000.1002c2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-269-0565
TELEFAX: 206-269-0563
TELEX:
INFORMATION FOR SEQ ID NO: 160:
SEQUENCE CHARACTERISTICS:
LENGTH: 541 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-997-362-160

Query Match 59.3%; Score 1574.5; DB 2; Length 541;
Best Local Similarity 61.3%; Pred. No. 5.7e-132;
Matches 323; Conservative 79; Mismatches 122; Indels 3; Gaps 2;

Qy 1 SKLIEYDEARRAMEVGMKDLADTVRVTLGPRGRHVLAFAFGPTVTNDGVTAAREIETL 60
Db 2 AKTIAYDEARGLERGLNALADAVKVTLPGRNVLEKKMGAPTITNDGVSIAKEIETL 61
Qy 61 EDPFEDLGAOLVKSVAATNDVAGDGTATITLAOLIKGRLVAAGVNPALGVGIGK 120
Db 62 EDPYKICAEVLKVAKTDDVAGDGTATITLAOLVREGIRNVAAGANPLGLKRGIEK 121
Qy 121 AADAVEALLASAPFVSKGTGIAQVATVSSRDEQIGDLVGEAMKNGHDSVVSKESSL 180
Db 122 AVEAVTOSLKSAREVERKEQISATTAISAGDTQIGELIAEAMDKVNGEVIYESMTF 181
Qy 181 GTELEFTGIGFHKGFSLAYEPTDNOQAVLEDAIILLHODKISSLDLPLEKVAAGT 240
Db 182 GLOLELTGMRDKGYSIGYFTDAEROEALEDPYILLVSSKSTVADLPLEKVIQA 241
Qy 241 GKPLIIVAEDEVEGELATLVNAIRKTLKAVAVKPGYGRDKAPLEDAVVTGSHVNP 300
Db 242 GKPLIIVAEDEVEGELSTLVNKRIGTFKSAVAKPFGDRKAMLOMALLTGQVYSE 301
Qy 301 DAGIVLEVEGLVGSARRVVSCKDVTIVDGGTAAEVANRANHLRAEIKSDSDMDRE 360
Db 302 RVGLSLETAADVSLIGQARKVVTKDETTIVEGSDSDAIAGVQAIRAEIENSDDYDRE 361
Qy 361 KLGERLAKLAGVAVIKGAATDTRALKERKESVEDAVAAKAAVEEGIVPGGASLIHQ 420
Db 362 KLGERLAKLAGVAVIKGAATEVELKERKHRIEDAVARNAAVEEGIVAGGVALL-QS 420
Qy 421 RKALTELASLTGDEVLDVDFSEALAPLPFIANAGLDGSSVYVKNKSELPAGHGLVN 480
Db 421 APALDDL--GLTGDEATGANIVRALSLAPLQIAFNGLEGEVVAEKSNLPAGHGLNAA 478
Qy 481 TLSYGDLAADVDPVKYTRSAVLNASSVARNVLTETVVDKPAKA 527
Db 479 TGEYEDLKAGVADPVKTRSAIQNASIALFLTTEAVVADKPEKA 525

RESULT 3
US-09-855-855-160
Sequence 160, Application US/09095855
Patent No. 6160093
GENERAL INFORMATION:
APPLICANT: Tan, Paul
APPLICANT: Visser, Elizabeth
APPLICANT: Skinner, Margot
APPLICANT: Prestidge, Ross
TITLE OF INVENTION: Compounds and Methods for
TREATMENT AND DIAGNOSIS OF MYCOBACTERIAL INFECTIONS
NUMBER OF SEQUENCES: 208
CORRESPONDENCE ADDRESS:
ADDRESSEE: Law Offices of Ann W. Speckman
STREET: 2601 Elliott Avenue, Suite 4185
CITY: Seattle
STATE: WA
COUNTRY: USA

```

? ZIP: 98121
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Diskette
? COMPUTER: IBM Compatible
? OPERATING SYSTEM: DOS
? SOFTWARE: FastSeq for Windows Version 2.0
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/09/095,855
? FILING DATE:
? CLASSIFICATION:
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: 08/705,347
? FILING DATE: 29-AUG-1996
? APPLICATION NUMBER: 08/873,970
? FILING DATE: 12-JUN-1997
? APPLICATION NUMBER: 08/997,362
? FILING DATE: 23-DEC-1997
? ATTORNEY/AGENT INFORMATION:
? NAME: Sleath, Janet
? REGISTRATION NUMBER: 37,007
? REFERENCE/DOCKET NUMBER: 11000.1002c3
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 206-269-0565
? TELEFAX: 206-269-0563
? TELEX:
? INFORMATION FOR SEQ ID NO: 160:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 541 amino acids
? TYPE: amino acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? US-09-095-855-160

```

```

Query Match          59.3%; Score 1574.5; DB 3; Length 541;
Best Local Similarity 61.3%; Pred. No. 5.7e-132;
Matches 323; Conservative 79; Mismatches 122; Indels 3; Gaps 2;

```

```

QY 1 SKLEIYDTRRRAMEVGMKLDATVRYTLCPRGRHVVYLAFAFGPIYVNDGVTVAARETEL 60
   :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 2 AKTIAYDEEARRGLEKRLNALDAVAVKVTLPGRNVVLEKKWGAPTTNDGVSIAKEIEL 61
QY 61 EDPEFDGALVSVAKTNDVAGDGTATITLAQALIKGLRLVAVGVPMALGVGICK 120
   :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 62 EDPEKIGALVEVAKKTIDVAGDGTITVLAQALVRGRLVAVAGANPLGKRIEIK 121
QY 121 ADAVFEALLASATPVSGKTGIAOVATVSSRDEQIGDLVEAMNKVGHDSVSKESSTL 180
   :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
b 122 AVEAVQSLKSAKEVETKROISATAISAGDQIGELINAEAMDKVNEGVIYVESNTF 181
QY 181 GTELEFTEGIGRHFGLSAFVTFDENOQAVLEBDALILHODKISSLPDLLPLEKYAGT 240
   :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 182 GLOLETEGMRFPKGYISGFVTDPEQEAVALDEDPYILLVSSKVSIVKDLLPLEKYIOA 241
QY 241 GKPLLLIAEVEGEALATLVVNAIRKTLKAVAVKGFEGDRRAKAFEDLAVVVGCHVNP 300
   :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 242 GKPLLLIAEVEGEALSTLVVNAIRKTFKSVAVKAFEGDRRAKAFEDLAVVVGCHVNP 301
QY 301 DAGIVLREVGLEVLGSAARVVVSKDDTVIVDGGGTAEAVANRANHRAETDKSDSDMDRE 360
   :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 302 RVLGSLTADVSLGQARKVVTIKDEFTTIVEGSDSDALINGRAVOIRAELEIENSDDSDRE 361
QY 361 KLGEBRLAKLAGVAIVKGAATDTALKERKESVEDAAVAAKAAVEGCIYVGGASLIHQ 420
   :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 362 KLOERLAKLAGVAIVKAGATEVELKERKHRIEDAVRNKKAVEREGIVAGGVALL-QS 420
QY 421 RKALTELRASTIGDEVLVGVNFSEALAPLFMTIAANAGLDSVVVNVKSLPGCHGNV 480
   :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 421 APPLDLDL--GLTDEATGAVIRVALSAPLKQAFNGLEPVAEVSALPAGHGLNAA 478
QY 481 TLISYGLDADGVTDVYKVTASAVLNASSVARMVLTETTVVNDKPRAKA 527
   :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 479 TGEYEDLLKAGVADPVKVTRSALONASTIALFLTTEAVVADKPEKA 525

```

```

RESULT 4
PCT-US94-06362-4
? Sequence 4, Application PC/TUS9406362
? GENERAL INFORMATION:
? APPLICANT:
? TITLE OF INVENTION: Stress Proteins and Uses Therefor
? NUMBER OF SEQUENCES: 4
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
? STREET: 2 Millitia Drive
? City: Lexington
? STATE: MA
? COUNTRY: USA
? ZIP: 02173
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: PatentIn Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: PCT/US94/06362
? FILING DATE: 06-JUN-1994
? CLASSIFICATION:
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/073,381
? FILING DATE: 04-JUN-1993
? ATTORNEY/AGENT INFORMATION:
? NAME: Granahan, Patricia
? REGISTRATION NUMBER: 32,227
? REFERENCE/DOCKET NUMBER: WHI88-08AP2A PCT
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (617) 861-6240
? INFORMATION FOR SEQ ID NO: 4:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 540 amino acids
? TYPE: amino acid
? TOPOLOGY: linear
? MOLECULE TYPE: protein
PCT-US94-06362-4

```

```

Query Match          59.2%; Score 1571.5; DB 4; Length 540;
Best Local Similarity 61.4%; Pred. No. 1e-131;
Matches 323; Conservative 73; Mismatches 127; Indels 3; Gaps 2;

```

```

QY 1 SKLEIYDTRRRAMEVGMKLDATVRYTLCPRGRHVVYLAFAFGPIYVNDGVTVAARETEL 60
   :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 2 AKTIAYDEEARRGLEKRLNALDAVAVKVTLPGRNVVLEKKWGAPTTNDGVSIAKEIEL 61
QY 61 EDPEFDGALVSVAKTNDVAGDGTATITLAQALIKGLRLVAVGVPMALGVGICK 120
   :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 62 EDPEKIGALVEVAKKTIDVAGDGTITVLAQALVRGRLVAVAGANPLGKRIEIK 121
QY 121 ADAVFEALLASATPVSGKTGIAOVATVSSRDEQIGDLVEAMNKVGHDSVSKESSTL 180
   :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 122 AVEKVTETLLKGAKEVETKROIAATAISAGDSIGELINAEAMDKVNEGVIYVESNTF 181
QY 181 GTELEFTEGIGRHFGLSAFVTFDENOQAVLEBDALILHODKISSLPDLLPLEKYAGT 240
   :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 182 GLOLETEGMRFPKGYISGFVTDPEQEAVALDEDPYILLVSSKVSIVKDLLPLEKYIOA 241
QY 241 GKPLLLIAEVEGEALATLVVNAIRKTLKAVAVKGFEGDRRAKAFEDLAVVVGCHVNP 300
   :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 242 GKPLLLIAEVEGEALSTLVVNAIRKTFKSVAVKAFEGDRRAKAFEDLAVVVGCHVNP 301
QY 301 DAGIVLREVGLEVLGSAARVVVSKDDTVIVDGGGTAEAVANRANHRAETDKSDSDMDRE 360
   :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 302 EYGLTFLENADLSLGRKARKVVTIKDEFTTIVEGAGDTDALINGRAVOIRAELEIENSDDSDRE 361
QY 361 KLGEBRLAKLAGVAIVKGAATDTALKERKESVEDAAVAAKAAVEGCIYVGGASLIHQ 420

```

Db 362 KLOERLAKAGVAVIRKGAATEVELKERKRIEDAVRNAAVEGIVAGGVTL-0A 420
Qy 421 RKALTELASLTGDEVLDVDFSEALAPLFWIANNAGLDGSVYVVKXSELPAGGLNVN 480
Db 421 APPTDELK--LEGDEATGANIVKALEAPLKQIAFNISGLEPGVAKRNPLPAGHGLNMQ 478
Qy 481 TLSYGDLAADGVDPVKYTRSAVLNASSVARNVLTETVYVVDKPAK 526
Db 479 TGVYEDLLAGVADPVKTRRSALQNAASLAGLFTTEAVNVADKPRK 524

RESULT 5
US-08-997-080-114
; Sequence 114, Application US/08997080
; Patent No. 5968524
; GENERAL INFORMATION:
; APPLICANT: WATSON, JAMES D.
; APPLICANT: TAN, PAUL L.J.
; TITLE OF INVENTION: METHODS AND COMPOUNDS FOR THE TREATMENT OF IMMUNOLOGICALLY-
; NUMBER OF SEQUENCES: 194
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Law Offices of Ann W. Speckman
; STREET: 2601 Elliott Avenue, Suite 4185
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98121
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/997,080
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Sleath, Janet
; REGISTRATION NUMBER: 37,007
; REFERENCE/DOCKET NUMBER: 11000.1007
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-269-0565
; TELEFAX: 206-269-0563
; TELEX:
; INFORMATION FOR SEQ ID NO: 114:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 523 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-997-080-114

Query Match 59.0%; Score 1566.5; DB 2; Length 523;
Best Local Similarity 61.3%; Pred. No. 2.8e-131;
Matches 321; Conservative 79; Mismatches 121; Indels 3; Gaps 2;

Qy 1 SKLEYDETARAMEVGMDKLADPVRYVTLGPRGRHVLAFAFGPTVNRDGVARETEL 60
Db 2 AKTAYIDEARRGLEGRNALADAKVTLGPRGRNVVLEKKMGAPTTINDGVSIAKEIEL 61
Qy 61 EDPFEDLGAQLVKVATNTDVAGDTTATILIAQALIKGRLIYAAGVNPVALGVGIGK 120
Db 62 EDPYEKIGAEIVKVKAKTDDVAGDGTATVIAQALVREGIRNVAAGANPLGKRGIEK 121
Qy 121 AADAVFALLASATPVSGKGTIAQAVATVSSRDEQIGDIVGEAMKNGVDVSVKESSTL 180
Db 122 AVEAVTOSLKSAREVEKTEQISATAISAGDTQIGELIAEAMDKVNGEIVTESWTF 181

Qy 181 GTELEFGIGFHKFSLAYFVTDFDNOAVLEDALILLHODKISSLPDLLPLEKKVAGT 240
Db 182 GLOLELTGMRFDKGTIGTYFTDAERQAVLEDPYILLVSSKSTVADLLPLEKVIOA 241
Qy 241 GKPLLIADVEGEALATLVNNAIRKTLKAVAVKPYFGDRRAFLLEDLAAYVGGHVPN 300
Db 242 GKPLLIADVEGEALSTLVNNAIRKTLKAVAVKPYFGDRRAFLLEDLAAYVGGHVPN 301
Qy 301 DAGYLVREVGLEVSARAVVSKDDYIYVDCGGTAELAEVANRANLIRAEIKSDSDMDRE 360
Db 302 RVGLSTETADVSLGQARKKVVYTDETTIVGSGSDAIAGRVQIRAEITNSDSDYDRE 361
Qy 361 KLGERTLAKGAVAVIRKGAATDTALKEKESVEDAVAAKAAVEGIVPGGASLIHQ 420
Db 362 KLOERLAKAGVAVIRKGAATEVELKERKRIEDAVRNAAVEGIVAGGVALL-0S 420
Qy 421 RKALTELASLTGDEVLDVDFSEALAPLFWIANNAGLDGSVYVVKXSELPAGHGLNVN 480
Db 421 APALDDL--GLTGDEATGANIVRALSPALQIAFNISGLEPGVAKVSNLPAGHGLNMA 478
Qy 481 TLSYGDLAADGVDPVKYTRSAVLNASSVARNVLTETVYVVDKPAK 524
Db 479 TGVYEDLLAGVADPVKTRRSALQNAASIALGLFTTEAVNVADKPRK 522

RESULT 6
US-08-997-362-114
; Sequence 114, Application US/08997362
; Patent No. 5985287
; GENERAL INFORMATION:
; APPLICANT: Tan, Paul
; APPLICANT: Hiyama, Jun
; APPLICANT: Visser, Elizabeth
; APPLICANT: Skinner, Margot
; APPLICANT: Prestidge, Ross
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR
; TREATMENT AND DIAGNOSIS OF MYCOBACTERIAL INFECTIONS
; NUMBER OF SEQUENCES: 194
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Law Offices of Ann W. Speckman
; STREET: 2601 Elliott Avenue, Suite 4185
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98121
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/997,362
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: U.S. Patent Application No. 5985287 08/873,970
; FILING DATE: June 12, 1997
; APPLICATION NUMBER: U.S. Patent Application No. 5985287 08/705,347
; FILING DATE: August 29, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Sleath, Janet
; REGISTRATION NUMBER: 37,007
; REFERENCE/DOCKET NUMBER: 11000.1002c2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-269-0565
; TELEFAX: 206-269-0563
; TELEX:
; INFORMATION FOR SEQ ID NO: 114:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 523 amino acids
; TYPE: amino acid

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-997-362-114

Query Match 59.0%; Score 1566.5; DB 2; Length 523;
Best Local Similarity 61.3%; Pred. No. 2.8e-131;
Matches 321; Conservative 79; Mismatches 121; Indels 3; Gaps 2;

QY 1 SKLEIYETARRAMEVGMKLDATVRLTGPGRHNVYLAKAFGPGTYTNDGVYARETEL 60
DB 2 AKTIAYDEEARRGLEGLNMLADAVKTLGPKGRNVYLEKKWGPATITNDGVSIKAEIEL 61
QY 61 EDPEDDGAOLVSVAKRTNDVAGDGTITATIIAOLIKGLRLVAGVNPVALGVIGK 120
DB 62 EDPKEKIGAEIVKAKRTDVGDTTATVLAQALVREGLRNVAAGANPLGKRIEIK 121
QY 121 AADAFBALLASATPVSGKTGIAOVATVSSRDEQIGDLVEANMKVGHDSVSKESSTL 180
DB 122 AVEAVTOSLKSAREVETKQISATTAISAGDTQIGELINEAMKVGNEGYITVEESNTF 181
QY 181 GTELEFTEGIGFHKGLSAYFVTDPNQOAVLEBALILHODKISSLPDLLPLEKVAQT 240
DB 182 GLOLETEGMRFPDQGYISGFVTDABRQEAULEDPYITLVSSKVTVDLLPLEKVIOA 241
QY 241 GKPLLIYAEDVEGALATLVNNAIRKTLKAVAKGPFGRKRAFLIEDLAVTGGHYNP 300
DB 242 GKPLLIYAEDVEGALSTLVNNAIRKTLKAVAKGPFGRKRAMLODMAITLGGQYVSE 301
QY 301 DAGIVREVGLEVLGSAARVVSCKDTVIYDGGGTAEAVANRANHLEAIKSDSDMDRE 360
DB 302 RVGLISTEADVSLGQARKVVTKDETTIVEGSDSDALIGRAVQAQIAELIENSDDRE 361
QY 361 KLGRLAKLAGVAVIKVGAATDTALKERKESVEDAAVAAKAAVEBIVGGASLIHOA 420
DB 362 KLGRLAKLAGVAVIKVGAATEVELKERKHRIEDAVRNKAAVEBIVAGGVALL-QS 420
QY 421 RKALTELASLTGDEVIGVDFSEALAPLFWIAANAGLDGSVVKNVSELPGHGLNVN 480
DB 421 APALDDL--GLTGDEATGANIVRVALSAPLKQIAFNGLPGVAAEVSNLPAGHGLNAA 478
QY 481 TLSYGDLAADGVIDPVKTRSAVLNASSVARMVLTETVVDKP 524
DB 479 TGEYEDLLKAGVADPVKTRSAVLNASSIALFLITTEAVVADKP 522

RESULT 7

US-09-095-855-114
Sequence 114, Application US/09095855
Patent No. 6160093
GENERAL INFORMATION:
APPLICANT: Tan, Paul
APPLICANT: Visser, Elizabeth
APPLICANT: Skinner, Margot
APPLICANT: Prestidge, Ross
TITLE OF INVENTION: Compounds and Methods for
TITLE OF INVENTION: Treatment and Diagnosis of Mycobacterial Infections
NUMBER OF SEQUENCES: 208
CORRESPONDENCE ADDRESS:
ADDRESSEE: Law Offices of Ann W. Speckman
STREET: 2601 Elliott Avenue, Suite 4185
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98121
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/095,855

FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/705,347
FILING DATE: 29-AUG-1996
APPLICATION NUMBER: 08/873,970
FILING DATE: 12-JUN-1997
APPLICATION NUMBER: 08/997,362
FILING DATE: 23-DEC-1997
ATTORNEY/AGENT INFORMATION:
NAME: Sleath, Janet
REGISTRATION NUMBER: 37,007
REFERENCE/DOCKET NUMBER: 11000.1002c3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-269-0565
TELEFAX: 206-269-0563
TELEX:
INFORMATION FOR SEQ ID NO: 114:
SEQUENCE CHARACTERISTICS:
LENGTH: 523 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-095-855-114

Query Match 59.0%; Score 1566.5; DB 3; Length 523;
Best Local Similarity 61.3%; Pred. No. 2.8e-131;
Matches 321; Conservative 79; Mismatches 121; Indels 3; Gaps 2;

QY 1 SKLEIYETARRAMEVGMKLDATVRLTGPGRHNVYLAKAFGPGTYTNDGVYARETEL 60
DB 2 AKTIAYDEEARRGLEGLNMLADAVKTLGPKGRNVYLEKKWGPATITNDGVSIKAEIEL 61
QY 61 EDPEDDGAOLVSVAKRTNDVAGDGTITATIIAOLIKGLRLVAGVNPVALGVIGK 120
DB 62 EDPKEKIGAEIVKAKRTDVGDTTATVLAQALVREGLRNVAAGANPLGKRIEIK 121
QY 121 AADAFBALLASATPVSGKTGIAOVATVSSRDEQIGDLVEANMKVGHDSVSKESSTL 180
DB 122 AVEAVTOSLKSAREVETKQISATTAISAGDTQIGELINEAMKVGNEGYITVEESNTF 181
QY 181 GTELEFTEGIGFHKGLSAYFVTDPNQOAVLEBALILHODKISSLPDLLPLEKVAQT 240
DB 182 GLOLETEGMRFPDQGYISGFVTDABRQEAULEDPYITLVSSKVTVDLLPLEKVIOA 241
QY 241 GKPLLIYAEDVEGALATLVNNAIRKTLKAVAKGPFGRKRAFLIEDLAVTGGHYNP 300
DB 242 GKPLLIYAEDVEGALSTLVNNAIRKTLKAVAKGPFGRKRAMLODMAITLGGQYVSE 301
QY 301 DAGIVREVGLEVLGSAARVVSCKDTVIYDGGGTAEAVANRANHLEAIKSDSDMDRE 360
DB 302 RVGLISTEADVSLGQARKVVTKDETTIVEGSDSDALIGRAVQAQIAELIENSDDRE 361
QY 361 KLGRLAKLAGVAVIKVGAATDTALKERKESVEDAAVAAKAAVEBIVGGASLIHOA 420
DB 362 KLGRLAKLAGVAVIKVGAATEVELKERKHRIEDAVRNKAAVEBIVAGGVALL-QS 420
QY 421 RKALTELASLTGDEVIGVDFSEALAPLFWIAANAGLDGSVVKNVSELPGHGLNVN 480
DB 421 APALDDL--GLTGDEATGANIVRVALSAPLKQIAFNGLPGVAAEVSNLPAGHGLNAA 478
QY 481 TLSYGDLAADGVIDPVKTRSAVLNASSVARMVLTETVVDKP 524
DB 479 TGEYEDLLKAGVADPVKTRSAVLNASSIALFLITTEAVVADKP 522

RESULT 8

US-08-467-822-34
Sequence 34, Application US/08467822
Patent No. 5843460
GENERAL INFORMATION:


```

OY      1 SKLIEYDFATRRAMVGVMDKLDADVRYTLGGRGHVYLAKAFGGPTVNDGVTAARETEL 60
Db      2 AKTIAYDEARRGLEGRGINSILADAVKTYLGLKGRNVYLEKKMGAPITINDGVSTAKELEL 61
OY      61 EDPEEDLCAQLVKSVAKRTNDVAGDGTTTATILAOALIKGSLRLVAAGVPVALGVGIGK 120
Db      62 EDPEYKIAEVLKVEYAKKTTDDVAGDGTITTAVALAOALKEKLRLVAAGANPLGLKRGIEK 121
OY      121 AADAVEALLSAPVPSCKGTGAQVATVSSRDEOIGDLVGEAMNKVGHDSVSVKESSTL 180
Db      122 AYDKYTELLNDAKAEVEYKEQDIAATAISAGDOSIGDLIAAMDKVNGEYITVEESNTF 181
OY      181 GTELEFTEGIGCFHKGFLSAFYVTFDFDNOAAVLEDALILLHODKISSLPDLLPLEKYAGT 240
Db      182 GLQLETFEGMGPRFDKGYISGYEFTDAREQEAULEEYPYILLVSSKSVTADLLPLEKYIQA 241
OY      241 GKPLTIVAEDEVEGELATLVNATRKTLKAAVAVGPGYGDGRKRAFLEDAVAVTGGHVPN 300
Db      242 GKSLTIIVDEVEGELSTLVVNNKIRGTYSVAAVAPGDDRRKAMLDOMALLTQAQVISE 301
OY      301 DAGIYLRVGEVLEVSARRVVVSCKDDPYIVVGGGTAEAVARAHRLRAEIDKSSDMDRE 360
Db      302 EYGLTLENTDLSLLGKAKKVMYTKDETTIVGAGGTDAIAGRVAGIIRIEINSDDVRE 361
OY      361 KLGEPLALAGVAVYTKGAATDTALAKRKRESVEDAAVAAKAAVEEGIVPEGGASLIHOA 420
Db      362 KLGEPLALAGVAAVIAKGAATEVELKERKHIRIDAVANNAKAAVEEGIVAGGVTLL-QA 420
OY      421 RKALTELRASTGLDGVLEVDVFESEALAPLFPMIANAAGLDGSVVNNKSELPAGHGVLVN 480
Db      421 APALDKLR--LTGDGATGANTIVKVALEAPLPQIAFNSGMEPGVAAEKRNLSVGHGLNAA 478
OY      481 TLSYGLDADGVIDPEKVTYSAVLNASSVARNVLTETTVVVDKPAK 526
Db      479 TGEYEDLLKAGVADPVKVTYSALONMAASIGLFLTTEVAVVADKPEK 524

RESULT 10
US-08-368-834-20
: Sequence 20, Application US/08368834
: Patent No. 5874405
: GENERAL INFORMATION:
: APPLICANT: Birnbaum, Gary
: APPLICANT: Kotlinek, Linda K.
: APPLICANT: Braun, Peter E.
: TITLE OF INVENTION: Heat Shock Protein Peptides That Share
: TITLE OF INVENTION: Sequences with Cyclic Nucleotide Phosphodiesterase and
: TITLE OF INVENTION: Methods for Modulating Autoimmune Central Nervous System
: NUMBER OF SEQUENCES: 20
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Merchant & Gould
: STREET: 3100 No. 5874405west Center
: CITY: Minneapolis
: STATE: Minnesota
: COUNTRY: USA
: ZIP: 55402
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/368,834
: FILING DATE: 16-DEC-1994
: CLASSIFICATION: 514
: ATTORNEY/AGENT INFORMATION:
: NAME: Kowalczyk, Katherine M.
: REGISTRATION NUMBER: 36,848
: REFERENCE/DOCKET NUMBER: 600.3030501

```

```

;      TELEPHONE: 612-332-5300
;      TELEFAX: 612-332-9081
;      INFORMATION FOR SEQ ID NO: 20:
;      SEQUENCE CHARACTERISTICS:
;          LENGTH: 540 amino acids
;          type: amino acid
;      STRANDEDNESS: single
;      TOPOLOGY: linear
;      MOLECULE TYPE: protein
;      OS-08-368-834-20

```

Best Local Similarity 60.1%; Pred.No. 1.9e-128;
Matches 316; Conservative 77; Mismatches 129; Indels 4; Gaps 3,

```

QY      1 SKLEVEYDSTARBRAMVGDMDKLADYVRLVLPGRGHRIVLAKAAGCGFTVTDGYTVAREIEL 60
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      2 AKTIAYDEARNGRLERGLNSLADAVKAVTLPGKGRNVVLEKKMGAPLTITNDGYSIAKEIEL 61
QY      61 EDPEEDLCAOLYKSVATRTNDVACGTTTATITLAAOLIKGILRLVYAAGVPAALGVGICK 120
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      62 EDPEKIKAEIYEVKAKRTDDVAGGTTTATVLAOLYKEGRIRNVAAGANPGLKRGIEK 121
QY      121 AADAVFEALILASATPVSGKTGIAQAVAFVSSRDEOIGTALVGEFAMNRYGHDVSYSVKESSL 180
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      122 AVDKYTELLILDAKAEYERKE-IAATPAISAGDOSIGDLIAEMDKVGNNGCVITVERSNTEF 180
QY      181 GTELEFTTEGIGFHKGFELSAYEFTVDFDNOQAVLEDALILLHODKISSLPDLPLELKVAGT 240
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      181 GLOLETFEGMGFMFDKGYISGYEVTDAERQEAVALFEFPYILLVSYSKVSTVKDLPLELEKVIQA 240
QY      241 GKPLITIVAEDEGEKALATLVNVAIKKTLKAVNVKRPYFEDRKKAPLEDAVATVYGGHVA 300
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      241 GKSLITIAEDVEGELSTLVNKKINGTERSAVNAKPGFEDRRKAMLODAILLGAOVISE 300
QY      301 DAGIYLRVGVGEVLCSARRVYVSKDDTVIVDGGTAEAVANRANHLRAEIDKSDSDMDRE 360
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      301 EVGLTLENTDLSILGKARKRVYMTKDETTIVBAGDPTDAIGRAVQIRTEIENSDDSDYRE 360
QY      361 KLGEFLALAGGAVIYKCATDTTLAKERKESVEDAVAAKAAVEEGIYPPGGCASLIHOA 420
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      361 KLGEFLALAGGAVAIKGAATEVLEKERRHIEIDAVRNAKAAVEEGIYAGGCVITLL-QA 419
QY      421 RKALTELKASLTGDDVNLVDVSEFSLAALPLFIANAGLGDGSVVYVYKSELPAHGCLYN 480
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      420 APALDKLK-LTGDDATGANIYKVALLEAPLKRIAINSGMEPEPVAAVEKYNANLSVGHOLMA 477
QY      481 TLVSGDLAADGVIDPVKYTRSAVLNASSAVARNVLTTEYVYVNDKPAK 526
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      478 TGEYEDLLKAGVADPVKRTSALQNAASIAGLFTTEAAVAVADKPEK 523

RESULT 11
FCT-US94-06362-3
: Sequence 3, Application PC/TUS9406362
: GENERAL INFORMATION:
: APPLICANT:
: APPLICANT:
: TITLE OF INVENTION: Stress Proteins and Uses Therefor
: NUMBER OF SEQUENCES: 4
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
: STREET: 2 Millitia Drive
: CITY: Lexington
: STATE: MA
: COUNTRY: USA
: ZIP: 02173
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS

```


Db 303 EEIGMELEKATLEDLGAKRVINKDPTTTIDGVEEAALQGRVAIQIREEATS DYDR 362
QY 360 EKIGERBLAKAGVAVITKGAATDTALKEKESVEDAAKAAVEGIVPGGASLIHQ 419
Db 363 EKLOERVAKLAGGAVAIKGAATEVMEKKAVEDALHATRAAEBEGVAGGVALIRV 422
QY 420 ARKALTELASLTGDEVLGVDVSEALAPLFIAANAGIDGSVVANKSELPAAGLNV 479
Db 423 ASR-LADLIHQ-NEDONVGKVALRAMEAPLRQIVLNCGEPSVAVNTYKGGDNGCYNA 480
QY 480 NTLISYDGLAADGVIDPVKTVRSAYLVNASSVARMVLTETVYVDDKP 524
Db 481 ATREYGNMIDMGILDPTKYTRSAIQYAASVAGLMTTECMVDLP 525

RESULT 13
US-08-467-822-31
; Sequence 31, Application US/08467822
; Patent No. 5843460
GENERAL INFORMATION:
APPLICANT: Labigne, Agnes
APPLICANT: Sauterdam, Sebastien
APPLICANT: Ferrero, Richard L.
APPLICANT: Thiberge, Jean-Michel
TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS AGAINST
TITLE OF INVENTION: HELICOBACTER INFECTION, POLYPEPTIDES FOR USE IN THE
TITLE OF INVENTION: COMPOSITIONS, AND NUCLEIC ACID SEQUENCES ENCODING SAID
NUMBER OF SEQUENCES: 44
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
ADDRESS: 1300 I Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/467,822
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/447,177
FILING DATE: 19-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/432,697
FILING DATE: 02-MAY-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Meyers, Kenneth J.
REGISTRATION NUMBER: 25,146
REFERENCE/DOCKET NUMBER: 03495, 0137-02000
TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4400
INFORMATION FOR SEQ. ID NO. 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 548 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-467-822-31

Query Match 48.8%; Score 1295.5; DB 2; Length 548;
Best Local Similarity 50.9%; Pred. No. 3.9e-107;

Matches 269; Conservative 105; Mismatches 152; Indels 3; Gaps 3;
QY 1 SKLIETDEARRAMEYGMKLDATVYVTLGPRGRHVLAKAREGPTVYNDGYTAREIEL 60
Db 2 AKELRFEGDARLQMLAGVNAADAVQVTPGRNVVLEKSYGAPVTRKDGVSVAKEIEF 61
QY 61 EDPEFELGALVSVATKNDVAGDCTTTATIAQLINGLURLVAAGVPAVACGIDK 120
Db 62 EHRFMNMGAMQAEVASKISDTRAGDGTATVATVARSILVEGHRAAVAGMPPMLKRGIDK 121
QY 121 AADAVEALLASATPVSGTGTIAQVAVTSR-DEQIGDVLGEAMNVGDSVVSKEST 179
Db 122 AVLAATVTKIQAMSKPKCKSKATIAQVCTISANSDEALGAIIAEMEVGAEVGTIVEDGNG 181
QY 180 LGTELEFTEGIGFHKGLSAYEVTDFDNOQAVLEDALILLHODKISSLPDLLPLEKVAG 239
Db 182 LENELVYVEGMQFDRGYISPYFINNOQNSCELEHFFILLVPRKVSIREMLSVLEGVAK 241
QY 240 TGRPLLIIVADVGEALATLVVNAIRKTLKAVAVKPGYGRDRKAPLEDLAVVGTGHVN 299
Db 242 SGRPLLIIVADVGEALATLVVNMGRIVKCAVKAPEGDRRKAMLDIATITRKQVTS 301
QY 300 PDAGIVLREYGLEVLGSARVVVSKDVTIVDGGTAEEAVANRANLRAEIDKSDSDMR 359
Db 302 EEIGKSLGEGATLEDLGASAKRIVTKENTTIIIOEGKATELNARIAQIRQMEETTS DYDR 361
QY 360 EKIGERBLAKAGVAVITKGAATDTALKEKESVEDAAKAAVEGIVPGGASLIHQ 419
Db 362 EKLOERVAKLAGGAVAIKGAATEVMEKKAVEDALHATRAAEBEGVAGGVALI-R 420
QY 420 ARKALTELASLTGDEVLGVDVSEALAPLFIAANAGIDGSVVANKSELPAAGLNV 479
Db 421 AQKALDSLQGD-NDDQMGINIRRAIESPMRQIVTNAGIEASVVKNAEHRDNVGFNA 479
QY 480 NTLISYDGLAADGVIDPVKTVRSAYLVNASSVARMVLTETVYVDDKPAKAE 528
Db 480 ATGEYDWMEMGILDPTKYTRMALQNAASVASLMLTTECMVADLPKKEE 528

RESULT 14
US-08-470-260-6
; Sequence 6, Application US/08470260
; Patent No. 6077706
GENERAL INFORMATION:
APPLICANT: Covacci, Antonello
APPLICANT: Bugnoli, Massimo
APPLICANT: Telford, John
APPLICANT: Macchia, Giovanni
APPLICANT: Rappoli, Rino
TITLE OF INVENTION: Helicobacter pylori Proteins Useful
NUMBER OF INVENTION: for Vaccines and Diagnostics
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chilton Corporation
STREET: 4560 Horton Street
CITY: Emeryville
STATE: California
COUNTRY: USA
ZIP: 94608-2916
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/470,260
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/256,848
FILING DATE: 21-OCT-1994
ATTORNEY/AGENT INFORMATION:
NAME: McClung, Barbara G.

```
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0316.001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 501-2708
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 546 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-470-260-6
```

```
Query Match          48.6%; Score 1289; DB 3; Length 546;
Best Local Similarity 50.1%; Pred. No. 1.5e-106;
Matches 265; Conservative 105; Mismatches 153; Indels 6; Gaps 3;
```

```
1 SKLIEYDETAARRAMEVGMKDLADTVRVTLGPRGRHVVLAARFGGPTVYNDGVYAREIEL 60
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2 AKEIKFSDSARNILFEGVRQLHDAVKVTMGPRGRNVLIQKSYGAPSTIKDGVSAKEIEL 61
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
61 EDPEDDGAQLVKRVAKTNDVAGDGTITTLAQLIKGRLVAAGVPVALGVGIGK 120
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
62 SCPVANNGAQLVKEVASTADAGDGTITATVLAYSIFKEGLRNTAGANPIEVKRGMDK 121
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
121 AADVFPALLASATPVSGKTGIAOVATVSSR-DEQIGDLVGEAMNKYGHDSVSKESST 179
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
122 AAELIINELKASKKVGKKEITTOVATISANSDHNIGKLADAMEKKGKDGVTVEEK 181
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
180 LGTELEFTEGIGFHKGLSAFVTVDFDNOQAVLEDAVLILHDKISSLPDLLPLEKVAG 239
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
182 IEDELDAVEGQDFRGYLSPEFVTAEMKMTAQLDNAYILLTDKKISSMKDILPLEKTMK 241
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
240 TGRKLLIIVAEDEVEGELATLVNNAIRKTLKAVAKGPFGRRAFLIEDLAIVYTGHHVN 299
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
242 ECKRLIIAIEDIEGELATLVNNAIRKTLKAVAKGPFGRRAFLIEDLAIVYTGHHVN 301
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
300 PDAGIVREVGLVGLSARRVNSKDPVTIVDGGGTAEAVANRANHRAEIDKSDSDMDR 359
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
302 EELGLSLENNEVEFLGAKGRIVIDKMTIYDGGHSDVDKRAQAQIKQTASTSDYDK 361
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
360 EKLGERLAKLAGVAIVIKGAATDTALKEKESVEDAAVAAKAAVEEGIVGGGASLIHQ 419
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
362 EKLGERLAKLSGVAIVIKGAASEVEEMKEKRDVDALSATKAVERGIVYGGGALIRA 421
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
420 ARKALTELRASLTGDEVLVGDFSEALAPLFWIAANAGLDGSVVNVKVSLEPLRAGHLNV 479
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
422 AQR---VHNLHDEKVEYGEIIRAIKAPLAQAIANAGYDGGVVNVNEVEKHEGFGFNA 477
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
480 NTLSTYGDLAADGVIDPVKTRSAVLNASSVARNVLTETGVVD-KPAKA 527
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
478 SNGKYVDMFKEGIIDPLKVERIALQNAVSVSLLITTEATVYHEIKEKA 526
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```
RESULT 15
US-08-471-491-6
; Sequence 6, Application US/08471491B
; Patent No. 6090611
; GENERAL INFORMATION:
; APPLICANT: Covacci, Antonello
; APPLICANT: Bugnoli, Massimo
; APPLICANT: Telford, John
; APPLICANT: Macchia, Giovanni
; APPLICANT: Rappuoli, Rino
; TITLE OF INVENTION: Helicobacter Pylori Proteins Useful For Vaccines And
; FILE REFERENCE: CHIR0044
; CURRENT APPLICATION NUMBER: US/08/471,491B
; CURRENT FILING DATE: 1995-06-06
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
```

```
; SEQ ID NO 6
; LENGTH: 546
; TYPE: PRF
; ORGANISM: Helicobacter pylori
US-08-471-491-6
```

```
Query Match          48.6%; Score 1289; DB 3; Length 546;
Best Local Similarity 50.1%; Pred. No. 1.5e-106;
Matches 265; Conservative 105; Mismatches 153; Indels 6; Gaps 3;
```

```
1 SKLIEYDETAARRAMEVGMKDLADTVRVTLGPRGRHVVLAARFGGPTVYNDGVYAREIEL 60
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2 AKEIKFSDSARNILFEGVRQLHDAVKVTMGPRGRNVLIQKSYGAPSTIKDGVSAKEIEL 61
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
61 EDPEDDGAQLVKRVAKTNDVAGDGTITTLAQLIKGRLVAAGVPVALGVGIGK 120
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
62 SCPVANNGAQLVKEVASTADAGDGTITATVLAYSIFKEGLRNTAGANPIEVKRGMDK 121
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
121 AADVFPALLASATPVSGKTGIAOVATVSSR-DEQIGDLVGEAMNKYGHDSVSKESST 179
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
122 AAELIINELKASKKVGKKEITTOVATISANSDHNIGKLADAMEKKGKDGVTVEEK 181
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
180 LGTELEFTEGIGFHKGLSAFVTVDFDNOQAVLEDAVLILHDKISSLPDLLPLEKVAG 239
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
182 IEDELDAVEGQDFRGYLSPEFVTAEMKMTAQLDNAYILLTDKKISSMKDILPLEKTMK 241
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
240 TGRKLLIIVAEDEVEGELATLVNNAIRKTLKAVAKGPFGRRAFLIEDLAIVYTGHHVN 299
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
242 ECKRLIIAIEDIEGELATLVNNAIRKTLKAVAKGPFGRRAFLIEDLAIVYTGHHVN 301
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
300 PDAGIVREVGLVGLSARRVNSKDPVTIVDGGGTAEAVANRANHRAEIDKSDSDMDR 359
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
302 EELGLSLENNEVEFLGAKGRIVIDKMTIYDGGHSDVDKRAQAQIKQTASTSDYDK 361
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
360 EKLGERLAKLAGVAIVIKGAATDTALKEKESVEDAAVAAKAAVEEGIVGGGASLIHQ 419
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
362 EKLGERLAKLSGVAIVIKGAASEVEEMKEKRDVDALSATKAVERGIVYGGGALIRA 421
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
420 ARKALTELRASLTGDEVLVGDFSEALAPLFWIAANAGLDGSVVNVKVSLEPLRAGHLNV 479
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
422 AQR---VHNLHDEKVEYGEIIRAIKAPLAQAIANAGYDGGVVNVNEVEKHEGFGFNA 477
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
480 NTLSTYGDLAADGVIDPVKTRSAVLNASSVARNVLTETGVVD-KPAKA 527
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
478 SNGKYVDMFKEGIIDPLKVERIALQNAVSVSLLITTEATVYHEIKEKA 526
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

Search completed: January 5, 2001, 14:27:16
Job time: 208 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 5, 2001, 14:23:23 ; Search time 62.98 Seconds
(without alignments)
292.097 Million cell updates/sec

Title: US-09-461-774-2

Perfect score: 2654

Sequence: 1 SKLEYDRTARRAMEVGMCK.....VVVDKRAKEDHHHGHAN 538

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 268485 seqs, 34193795 residues

otal number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq.36.*
1: /SIDSI/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq/AA2000.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2596	97.8	539	20	Y23914	Amino acid sequenc
2	2223	83.8	537	20	Y23912	Amino acid sequenc
3	1643.5	61.9	540	13	R22363	GroEL-1 protein.
4	1574.5	59.3	540	9	P81351	Sequence of Mycob
5	1574.5	59.3	540	16	R81610	Mycobacterium tube
6	1574.5	59.3	540	19	W44702	Mycobacterium tube
7	1574.5	59.3	540	20	Y23911	Mycobacterium tube
8	1574.5	59.3	541	20	Y14909	Amino acid sequenc
9	1574.5	59.3	560	9	P80215	Amino acid sequenc
10	1572.5	59.3	540	18	W32100	Sequence of Mycob
11	1566.5	59.0	523	19	W60144	Mycobacteria sp. h
12	1566.5	59.0	523	20	Y14891	M. vaccae antigen
						Amino acid sequenc

13	1555.5	58.6	545	20	Y23930	Consensus amino aci
14	1554.5	58.6	544	18	W32099	Mycobacteria sp. h
15	1552	58.5	572	11	R04716	Amino acid sequenc
16	1552	58.5	573	16	R64766	M. tuberculosis 65
17	1543.5	58.2	541	16	R67384	M. tuberculosis 65
18	1543.5	58.2	541	20	Y23910	M. leprae GroEL 1 g
19	1543.5	58.2	588	9	P80364	Amino acid sequenc
20	1540.5	58.0	541	20	Y23913	M. leprae 65kd ant
21	1525	57.5	539	20	Y23919	Amino acid sequenc
22	1520.5	57.3	548	17	R94368	Amino acid sequenc
23	1507	56.8	573	11	R04715	Brevibacterium fla
24	1503	56.6	573	16	R64765	Amino acid sequenc
25	1427.5	53.8	545	20	Y23904	M. leprae 65 kDa p
26	1399	52.7	539	20	Y23906	Streptococcus pyog
27	1398.5	52.7	544	20	Y23905	Amino acid sequenc
28	1393	52.5	541	20	Y23917	Amino acid sequenc
29	1375.5	51.8	541	20	Y23902	Streptococcus pneu
30	1369	51.6	552	20	Y23918	Amino acid sequenc
31	1355.5	51.1	548	18	W16678	Lawsonia intracell
32	1351	50.9	440	13	R22362	GroEL-1 protein pa
33	1318	49.7	539	20	Y23916	Amino acid sequenc
34	1316.5	49.6	544	21	Y75747	Neisseria meningit
35	1314.5	49.5	545	20	Y23915	Amino acid sequenc
36	1307.5	49.3	548	19	W65070	E. coli GroEL prot
37	1307.5	49.3	548	20	W95001	W09902989 Seq ID 1
38	1306.5	49.2	544	12	R13337	HyPB protein. Chl
39	1305.5	49.2	548	19	W61378	GroEL N-terminal r
40	1301.5	49.0	544	20	Y23903	Streptococcus pyog
41	1300.5	49.0	549	20	Y23909	Amino acid sequenc
42	1299.5	49.0	545	20	Y23901	Streptococcus pneu
43	1299.5	49.0	548	16	R67382	E. coli GroEL gene
44	1289	48.6	546	14	R11661	Hsp. Hsp60.
45	1289	48.6	546	14	R41200	Hsp. Helicobacter

ALIGNMENTS

RESULT 1	
Y23914	Y23914 standard; Protein; 539 AA.
ID	
XX	Y23914;
AC	22-SEP-1999 (first entry)
XX	
DT	Amino acid sequence of a heat shock protein.
XX	
DE	Heat shock protein; Hsp; immune response; immunological carrier;
XX	cancer control; tumour; sarcoma; cancer; gene therapy.
KW	
XX	
OS	Mycobacterium tuberculosis.
XX	
PN	W09935270-A1.
XX	
PD	15-JUL-1999.
XX	
PE	29-DEC-1998; 98WO-CA01203.
XX	
PR	31-DEC-1997; 97US-0001737.
XX	
PA	(STRE-) STRESSGEN BIOTECHNOLOGIES CORP.
XX	
PI	Mizzen L, Wisniewski J;
XX	
DR	WPI; 1999-430397/36.
XX	
PT	New nucleic acid encoding heat shock protein-60 from Streptococcus,
PT	useful in vaccines, as carriers for other immunogens, as anticancer
XX	agents and for diagnosis
PS	Disclosure; Fig 10A-E; 176pp; English.
XX	

CC Y23905-30 represent heat shock proteins (Hsps). The specification
CC describes Streptococcal Hsps, designated Hsp60. These proteins, their
CC fragments, variants and fusion proteins, are used to elicit or enhance
CC an immune response against Streptococcus, and to elicit a similar
CC response to a target antigen fused to the protein. Unlike other
CC immunological carriers, Hsp60 proteins are not immunosuppressive so
CC provide an increased response to any conjugated or fused antigen. Also,
CC where used for cancer control, they lack the side effects associated
CC with endotoxins. They can also be used to detect specific antibodies
CC and in treatment or prevention of tumours (e.g. sarcoma or cancers of
CC breast, ovary, prostate, lung, pancreas or liver). The Hsp60
CC polynucleotide is used for recombinant production of the protein, as
CC a source of primers and probes for detecting streptococci in standard
CC hybridization/amplification assays, and therapeutically in gene
CC therapy vectors.

XX
XX Sequence 539 AA:

Query Match 97.8%; Score 2596; DB 20; Length 539;
Best Local Similarity 98.0%; Pred. No. 3.2e-171;
Matches 527; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

QY 1 SKLIEYDTARRAMEVGMKDLADIVRVTLGPRGRHVLAKAFGPTVNDGVVARETEL 60
|||
DB 2 sklieydetarramevgmkladivrvtlgprgrhvlakafgptvndgvvaretel 61

QY 61 EDPEDIGAOLVKSVAATKTNVAGDGTATITLQAALIKGRLVAAGVNPVALGVGIGK 120
|||
DB 62 eppedigaoolvksvatktnvagdgtatitlqaalikgrlvaagvnpvalgvigk 121

QY 121 AADAVFALLASATPVSGKTGIAOVATVSSRDEQIGDLVGEAMKVKHDSVSVKESSTL 180
|||
DB 122 aadvfllasatpvsgktgiaovatvssrdeqigdlvgeamkvkhdsvsvkesstl 181

QY 181 GTELEFTGIGFHNKGFSAFVTPDFDNOQAVLEBALITLHODKISSLPDLPLEKVAAGT 240
|||
DB 182 gteleftgigfhnkgsafvtpdfdnqavlebalitlhodkisslpdlplekvagt 241

QY 241 GKPLLIIVAEDEGEALATLVNNAIRKTLKAVANGPVGDRKRAFLEDAVVTGGHVVNP 300
|||
DB 242 gkplliivaedegaalatlvnnairktlkavangpvgdrkrafledavvtggvnp 301

QY 301 DAGIVLREVGLVLSGARVVVSKDDTVIVDGGTAFAVANRANHLRAEIDKSDSDMDRE 360
|||
DB 302 dagvllrevglvlsgarvvvskddtviivdggtafaavanranhlraeidksdmdre 361

QY 361 KLGERLAKLAGVAVIYKGAATDPALEKREKSEVEDAVAAKAAVEEGIVPGGASLIHQ 420
|||
DB 362 klgerlaklagvavikygaatdpalekrekesvedavaakaaaveegivpggaslihq 421

QY 421 RKALTELRASTLGDVGLVDFSELAAPLFIWANAAGLDSVVVNKSELPAGHGLNVN 480
|||
DB 422 rkaltelrastlgdvglvdfsealaplfiiwanaagldsvvvnkselepaghglvnn 481

QY 481 TLSTGDLAADGVIDPVKTRSAVINASSVAMVLTFTTVVVDKPAKADHDHGHGH 538
|||
DB 482 tlstgdlaadgvidpvktrsavinassvamlftttvvvdkpaakedhghgh 539

RESULT 2
ID Y23912
XX Y23912 standard; Protein: 537 AA.
AC Y23912;
XX
XX 22-SEP-1999 (first entry)
DE Amino acid sequence of a heat shock protein.
XX
XX Heat shock protein; Hsp; immune response; immunological carrier;
KM cancer control; tumour; sarcoma; cancer; gene therapy.
XX

OS Mycobacterium leprae.
PN WO935270-A1.
XX
XX 15-JUL-1999.
XX
XX 29-DEC-1998; 98WO-CA01203.
XX
XX 31-DEC-1997; 97US-0001737.
XX
XX (STRE-) STRESSGEN BIOTECHNOLOGIES CORP.
XX
XX Mlzen L, Wisniewski J;
XX
XX WPI; 1999-430397/36.
XX
XX New nucleic acid encoding heat shock protein-60 from Streptococcus,
XX useful in vaccines, as carriers for other immunogens, as anticancer
XX agents and for diagnosis

PS Disclosure; Fig 10A-E; 176pp; English.

CC Y23905-30 represent heat shock proteins (Hsps). The specification
CC describes Streptococcal Hsps, designated Hsp60. These proteins, their
CC fragments, variants and fusion proteins, are used to elicit or enhance
CC an immune response against Streptococcus, and to elicit a similar
CC response to a target antigen fused to the protein. Unlike other
CC immunological carriers, Hsp60 proteins are not immunosuppressive so
CC provide an increased response to any conjugated or fused antigen. Also,
CC where used for cancer control, they lack the side effects associated
CC with endotoxins. They can also be used to detect specific antibodies
CC and in treatment or prevention of tumours (e.g. sarcoma or cancers of
CC breast, ovary, prostate, lung, pancreas or liver). The Hsp60
CC polynucleotide is used for recombinant production of the protein, as
CC a source of primers and probes for detecting streptococci in standard
CC hybridization/amplification assays, and therapeutically in gene
CC therapy vectors.

XX
XX Sequence 537 AA:

Query Match 83.8%; Score 2223; DB 20; Length 537;
Best Local Similarity 82.3%; Pred. No. 1.5e-145;
Matches 443; Conservative 48; Mismatches 45; Indels 2; Gaps 2;

QY 1 SKLIEYDTARRAMEVGMKDLADIVRVTLGPRGRHVLAKAFGPTVNDGVVARETEL 60
|||
DB 2 sklieydetarramevgmkladivrvtlgprgrhvlakafgptvndgvvaretel 61

QY 61 EDPEDIGAOLVKSVAATKTNVAGDGTATITLQAALIKGRLVAAGVNPVALGVGIGK 120
|||
DB 62 eppedigaoolvksvatktnvagdgtatitlqaalikgrlvaagvnpvalgvigk 121

QY 121 AADAVFALLASATPVSGKTGIAOVATVSSRDEQIGDLVGEAMKVKHDSVSVKESSTL 180
|||
DB 122 aadvfllasatpvsgktgiaovatvssrdeqigdlvgeamkvkhdsvsvkesstl 181

QY 181 GTELEFTGIGFHNKGFSAFVTPDFDNOQAVLEBALITLHODKISSLPDLPLEKVAAGT 240
|||
DB 182 gteleftgigfhnkgsafvtpdfdnqavlebalitlhodkisslpdlplekvagt 241

QY 241 GKPLLIIVAEDEGEALATLVNNAIRKTLKAVANGPVGDRKRAFLEDAVVTGGHVVNP 300
|||
DB 242 gkplliivaedegaalatlvnnairktlkavangpvgdrkrafledavvtggvnp 301

QY 301 DAGIVLREVGLVLSGARVVVSKDDTVIVDGGTAFAVANRANHLRAEIDKSDSDMDRE 360
|||
DB 302 etgvlrevglvlsgarvvvskddtviivdggtafaavanranhlraeidksdmdre 361

QY 361 KLGERLAKLAGVAVIYKGAATDPALEKREKSEVEDAVAAKAAVEEGIVPGGASLIHQ 420
|||
DB 362 klgerlaklagvavikygaatdpalekrekesvedavaakaaaveegivpggaslihq 420


```

OY 421 RKALTELRLASLTGDEVLDVDFSEALAPLEFWIAANAGLDGVSVMKSELPAHGHLNV 480
    ||:||||| ||:|||| ||:|||| ||:|||| ||:|||| ||:|||| ||:|||| ||:||||
DB 421 gaalkqrltsltgdealgivfleaalkaplywlatmagldgavvdkvsglpgnlnas 480
OY 481 TSLGYDLAADGVIDPVKVTBSAVLNASVAVMLTTEVTVDKPARAKEDHHHGHAN 538
    ||:|||| ||:|||| ||:|||| ||:|||| ||:|||| ||:|||| ||:|||| ||:||||
DB 481 tlgysgdlvadgvdpvkvtbsavlnaasvarmltletavvdkpakteehnd-bhgnah 537

RESULT 3
R22363
ID R22363 standard; Protein: 540 AA.
XX
AC R22363;
XX
DT 29-JUL-1992 (first entry)
XX
DE GroEL-1 protein.
XX
KW Heat shock protein; groEL-1 gene; HSP18.
XX
OS Streptomyces albus.
XX
FH Key 1..170 Location/Qualifiers
FT Protein /note="HSP18 heat shock protein"
FT
XX
PN WO9204452-A.
XX
PD 19-MAR-1992.
XX
PF 03-SEP-1991; 91WO-FR00701.
XX
PR 10-SEP-1990; 90FR-0011186.
XX
PA (INSP) INST PASTEUR.
XX
PI Mazodier P, Guglielmi G;
XX
DR WPI; 1992-114358/14.
DR NSDB; Q22482.
XX
PT Recombinant DNA contg. heat inducible promoter and heterologous
PT gene - also vectors, transformed cells and new heat shock
PT proteins of Streptococcus albus
XX
PS Claim 18; Fig 6; 50pp; French.
XX
CC The amino acid sequence is that of the 56 kDa GroEL-1 protein which
CC is encoded by the groEL-1 gene, it contains the 18 kDa HSP18 heat
CC shock protein. See also R22362 and R22364.
XX
SQ Sequence 540 AA;

Query Match 61.9%; Score 1643.5; DB 13; Length 540;
Best Local Similarity 61.0%; Pred. No. 1.2e-105;
Matches 330; Conservative 87; Mismatches 119; Indels 5; Gaps 4;
OY 1 SKLIEYDETRARAMEVMDKLADTVRVTLGPRGRHVLAFAFGPVTNDGVAREIEL 60
    ||:|||| ||:|||| ||:|||| ||:|||| ||:|||| ||:|||| ||:|||| ||:||||
DB 2 aklkkdeedarrlergvgadltvkvitgpkrynvidkfkfapcltndgvtiarevec 61
OY 61 EDPFEDLGAQLVKSVAFTKINDVAGDGTATTIAQALIKGLRLVAGVNPVALGVGICK 120
    ||:|||| ||:|||| ||:|||| ||:|||| ||:|||| ||:|||| ||:|||| ||:||||
DB 62 ddqyenlgqqlvkeavtkndiagdgltlatvialagelivreglnvaagaapaalkkxida 121
OY 121 AADAVEALLASATPVSGKTGIAOVATVSSRDQIGDLVGEANMKYGHDSVYVKESSTL 180
    ||:|||| ||:|||| ||:|||| ||:|||| ||:|||| ||:|||| ||:|||| ||:||||
DB 122 aavaavaealdtarpidkdsdaaavaalsqdkqvgellaaeandkvkgdgvltveesntf 181
OY 181 GTLEFTEGIGFHKGFLSAFVFTDFDNOQAVLEDALELHODKISSLPDLPLEKVV-A 238
    ||:|||| ||:|||| ||:|||| ||:|||| ||:|||| ||:|||| ||:|||| ||:||||

```

```

DB 182 gvaldfteegmatfdkgyLsppymrvdgermeavldddy11lhqgkigsigdl1p1lekviga 241
OY 239 GIGKPLLIYAEDVEGALATLVNNAIRKTLKAVAKGPFYFGDRRAKFLLEDLAVTGHNV 298
    ||:|||| ||:|||| ||:|||| ||:|||| ||:|||| ||:|||| ||:|||| ||:||||
DB 242 gsskpl11laedvegealstlvvnlrgtlnaavakapfgdtrkam1gdmaltlgaev1 301
OY 299 NPAGIVLREVGLEVIGSARRVVSFKDPRVINYDGGTAEAVNRAHMLRAETDKSDSDMD 358
    ||:|||| ||:|||| ||:|||| ||:|||| ||:|||| ||:|||| ||:|||| ||:||||
DB 302 aeevg1klldqagldvlgtravtlvctkddtclvdggnaedvqgvaqlkae1estdsdwd 361
OY 359 REKLGERLAKLAGVAVIRKGAATDTALKERKESVEDAAVAAKAEIVPGCASLIH 418
    ||:|||| ||:|||| ||:|||| ||:|||| ||:|||| ||:|||| ||:|||| ||:||||
DB 362 reklgerlalklggyvctirvgaatevelkerkhrleada1sattraaveg1vsggsalvh 421
OY 419 QARKALTELRLASLTGDEVLDVDFSEALAPLEFWIAANAGLDGVSVMKSELPAHGHLN 478
    ||:|||| ||:|||| ||:|||| ||:|||| ||:|||| ||:|||| ||:|||| ||:||||
DB 422 -avkvlvd-1lgrtgdgaatcgvavavrraavep1rwlaenag1egylvltkvaeldkqgfn 479
OY 479 VNPLSYGDLAADGVIDPVKVTBSAVLNASVAVMLTTEVTVDKPARAKAE-DHHHGHAN 537
    ||:|||| ||:|||| ||:|||| ||:|||| ||:|||| ||:|||| ||:|||| ||:||||
DB 480 aalgeygdlvkagvldpvkvtrsalenaas1asl1lttelvvekpaeepag1hghns 539
OY 538 H 538
DB 540 h 540

RESULT 4
ID P81351
XX P81351 standard; protein: 540 AA.
XX
AC P81351;
XX
DT 29-DEC-1990 (first entry)
XX
DE Sequence of Mycobacterium tuberculosis 540 AA residue protein.
XX
KW Diagnosis; assay; M.bovis; vaccine.
XX
OS Mycobacterium tuberculosis.
XX
PN WO8806591-A.
XX
PD 07-SEP-1988.
XX
PF 25-FEB-1988; 88WO-US00598.
XX
PR 24-FEB-1988; 88US-0159667.
XX
PA (SCRI-) SCRIPPS CLINIC & RE.
XX
PI Shinnick T, Houghten R;
XX
DR WPI; 1988-271136/38.
DR N-PSDB; N81768.
XX
PT Recombinant mycobacterial peptide(s) -
PT used in assays for diagnosis of infection, for producing
PT vaccines and for producing antibodies
XX
PS Disclosure; Fig 2a-2d; 116pp; English.
XX
CC An isolated DNA molecule that consists essentially of the nucleotide
CC sequence that corresponds to the sequence represented by position 3950 to
CC about 2390 and from position 3948 through position 2398 of N81768 is
CC claimed. Also claimed is a peptide sequence that consists of a 5-40 AA
CC residue sequence that corresponds to a sequence of the 540 AA residue
CC protein (P81351) or the 517 AA residue protein (P81868) coded for by the
CC DNA sequence. The proteins can be used for determining previous
CC immunological exposure of a mammal to M.tuberculosis or M.bovis and
CC for producing a vaccine.
XX
SQ Sequence 540 AA;

```


XX	W44702.
AC	
XX	22-JUN-1998 (first entry)
DT	
XX	Mycobacterium tuberculosis 65 kDa heat shock protein (Hsp65).
DE	
XX	
KW	Heat shock protein; Mt Hsp65; autoimmune disease; immunotherapy;
KW	gene therapy; rheumatoid arthritis; multiple sclerosis.
XX	
OS	Mycobacterium tuberculosis.
XX	
PN	W09746253-A2.
XX	
PD	11-DEC-1997.
XX	
XX	03-JUN-1997; 97WO-US09427.
XX	
PR	03-JUN-1997; 97US-0019100.
PR	03-JUN-1996; 96US-0019100.
XX	
PA	(AURA-) AURAGEN INC.
PI	Haynes JR, Prayaga SK, Ramshaw IA;
XX	
DR	WPI, 1998-041892/04.
DR	N-PSDB, V05708.
XX	
PT	Treatment of autoimmune diseases - by administering
PT	autoantigen-coated particles or autoantigen-encoding nucleic acid
PT	construct
PS	
XX	Example 2; Page 55-58; 72pp; English.
XX	
CC	This protein comprises the 65 kDa heat shock protein, Mt Hsp65, of
CC	Mycobacterium tuberculosis. This protein cross-reacts with a
CC	component of articular cartilage, human Hsp60, that is up-regulated
CC	in the joints of arthritic patients. The Mt Hsp65 gene sequence
CC	(see V05708) is also provided. A claimed method for treating or
CC	preventing an autoimmune disease in a mammal comprises: (a)
CC	providing a particle coated with an antigen against which an immune
CC	response is mounted in the autoimmune disease; (b) delivering the
CC	particle into the recipient cell of the mammal; and (c) repeating the
CC	step (b) until either a reduction in a cytotoxic immune response or
CC	a desensitizing immune response is induced in the mammal.
CC	Alternatively, step (a) comprises providing a nucleic acid construct
CC	comprising a coding sequence for the antigen, operably linked to
CC	control elements such that the coding sequence can be transcribed
CC	and translated in a recipient cell, and delivering the construct to
CC	the recipient cell using a gene gun. The antigen of step (a) is
CC	selected from collagen, Mt Hsp65, myelin basic protein, myelin
CC	oligodendrocyte glycoprotein, proteolipid protein, and epitopes
CC	thereof. These antigens mitigate cytotoxic responses and elicit
CC	antigen desensitisation. The method is used especially for treating
CC	rheumatoid arthritis or multiple sclerosis. It represents a novel
CC	use for the previously known Mt Hsp65.
CC	
XX	Sequence 540 AA;
SQ	
XX	
Query Match	59.3%; Score 1574.5; DB 19; Length 540;
Best Local Similarity	61.2%; Pred. NO. 6.5e-101;
Matches 322; Conservative	75; Mismatches 126; Indels 3; Gaps 2;
QY	1 SKLEYDETARAMEVGMKDLADRYRLTGPGRHVLAKAFGCGFTVNDGTVTARETEL 60
DB	2 AKLAYDEEARTRGLERGINALADAKVKVLPGYGRVIVLEKKWAGPCLTINDYSIAKEIEL 61
QY	61 EDPEEDLAAQVLKVSATNTNVAGDGTATLAAQALIKGRLATVAAGVNPALCVGIGK 120
DB	62 EDPEYKIGAEIVKEAKKTDVAGDGTATLAAQALIKGRLATVREGLINVAAGANPIGLIRGLEK 121
QY	121 AADAVFEALLASATPVSGKTIAGQAVATVSSRDEQIGDLVGEAMNKVGHDSVYSKESSTL 180

Db	122	avekvtecllkgakevteclqciaataisaqdslygdllaemdkvngnevytveesnltf	181
Qy	181	GTELEFTEGIGFHHGFLSAFVTDFFNQOAVLEDAILLHODKISSLPDLPLEKYAGT	240
Db	182	gqlqetlegmfdkgytssgyfvdpergeavledpylllvsskstvxdllpillekylga	241
Qy	241	GKPLLIYAEDVEGALATLVNAINRKTLLKAAVAKGPFESGDRKAFLEDLAAVVGCHVNP	300
Db	242	gkpllliaedvegealstlvnklrtgfcfskavkpgpfgdrtkamlqdmalltggqvise	301
Qy	301	DAGIVLENGVEGVSGARRRVYVKSDDTVYVDGSGTAENANANLRLREIDKSSDMDRE	360
Db	302	evgiltlenadslilgkarkkvtekldeitlvegagtdaagvraqiqelensdsdydre	361
Qy	361	KLGSRLAKTLAGVAVIVKGAATDTALKERRESVEDAVAAAKAAVEGIVPCGASLTHQA	420
Db	362	klgerklaklagvavikagatetevelkerkhrledavrnakaaveeglvaggyvtll-qe	420
Qy	421	RKALTELRASITGDEVLGVDFSEALADPLFWIAANAGIDGSVVVKNVKSLEPACHGLNVN	480
Db	421	aprldealk--legdeatganikvvalleaaplkqiafnsjlepgyvaekrnlpadghlnag	478
Qy	481	TLISGDLAADGVIDPVKFTSSAFLNNSSVARWVLTTEYVYVNDVPAK	526
Db	479	tygvdiilaagvadpvkvrtsalqnaaslaagllfteavvadkpek	524
RESULT	7		
Y23911			
ID	Y23911	standard; Protein: 540 AA.	
AC	Y23911;		
XX			
XX	22-SEP-1999	(first entry)	
DT			
XX			
DE		Amino acid sequence of a heat shock protein.	
XX			
KW		Heat shock protein; Hsp; immune response; immunological carrier;	
KW		cancer control; tumour; sarcoma; cancer; gene therapy.	
XX			
OS		Mycobacterium bovis.	
XX			
PN		W09J935270-A1.	
PD		15-JUL-1999.	
XX			
PF		29-DEC-1998; 98WO-CA01203.	
XX			
PR		31-DEC-1997; 97US-0001737.	
XX			
PA		(STRE-) STRESSGEN BIOTECHNOLOGIES CORP.	
XX			
PI		Mitzen L, Wisniewski J;	
XX			
DR		WPI: 1999-430397/36.	
PT		New nucleic acid encoding heat shock protein-60 from Streptococcus,	
PT		useful in vaccines, as carriers for other immunogens, as anticancer	
XX		agents and for diagnosis	
XX			
PS		Disclosure; Flg 10A-E; 176pp; English.	
XX			
CC		Y23905-30 represent heat shock proteins (Hsps). The specification	
CC		describes Streptococcal Hsps, designated Hsp60. These proteins, their	
CC		fragments, variants and fusion proteins, are used to elicit or enhance	
CC		an immune response against Streptococcus, and to elicit a similar	
CC		response to a target antigen fused to the protein. Unlike other	
CC		immunological carriers, Hsp60 proteins are not immunosuppressive so	
CC		provide an increased response to any conjugated or fused antigen. Also	
CC		where used for cancer control, they lack the side effects associated	
CC		with endotoxins. They can also be used to detect specific antibodies	
CC		and in treatment or prevention of tumours (e.g. sarcoma or cancers of	

breast, ovary, prostate, lung, pancreas or liver). The Hsp60 polynucleotide is used for recombinant production of the protein, as a source of primers and probes for detecting streptococci in standard hybridization/amplification assays, and therapeutically in gene therapy vectors.

Sequence 540 AA;

Query Match 59.3%; Score 1574.5; DB 20; Length 540;
Best Local Similarity 61.2%; Pred. No. 6.5e-101;
Matches 322; Conservative 75; Mismatches 126; Indels 3; Gaps 2;

1 SKLEIYDFTARAMEVGDKLDVRYTLGPRGRHVLAFAFGPTVNDGTVARETEL 60
2 AKLIAYDEAATRGIERGNAIADAVKVLIPKGRNVLEKKWGAPELNDGVSIAKEI 61
61 EDPEDLGAQLVKSVAATKNDVAGDGTATILAOALIKGLRLVAAGVNFALGVGIGK 120
62 EDPYKIGAEIVKEVAKKTDVAGDGTATLVLGALVREGLRVAAGANPLIGIKYIEK 121
121 AADAVFELASATPVSKGTIAOVATVSSRDEQIGDLVGEAMKVGHDVSVSKESSTL 180
122 AVEKVTCELIGAKVEVEKQIAATAISAGDSIGDLIAEAMDKVNGEYITVEESNLF 181
181 GTELEFTEGIGFHKGFLSAYFVTDPDNOQAVLEDAILILHODKISSLPDLPLEKXVAGT 240
182 GIGLEIPEGMTFIDKYISGYITVDPERGEAVLEDPYILIVSKSVTVKDLIPLEKVI 241
241 GKPLIIAEDVEGALATLVNNAIRKTLKAVAVGKPYGDRRKAFLEDLAVVTGHHVNP 300
242 GKPLIIAEDVEGALATLVNNAIRKTLKAVAVGKPYGDRRKAFLEDLAVVTGHHVNP 301
301 DAGIVLRVGEIVGSAARVYVSKDDIYVDDGGTAEEVANRANHLAEIDKSNDMDRE 360
302 EGVLTLEADISLIGKAKVVKVETDEIVGEGDQDAIGRAVQIQLIENSADYDRE 361
361 KLGRIELAKGAVIVIKGATDPTALKERKESVEDAAVAAKAEVIGPGASLIHQ 420
362 KIGERLAKGAVIVIKGATDPTALKERKESVEDAAVAAKAEVIGPGASLIHQ 420
421 RKALTELASLTGDEVLCVDFSEALAPLFWIANAGLDGSSVVNVKSELPAHGILNVN 480
421 AKPLIETASLTGDEVLCVDFSEALAPLFWIANAGLDGSSVVNVKSELPAHGILNVN 480
481 TLSGDLADGVIDPVKTRBSAVLNASSAVRNVLTFTRVVVDKPAK 526
479 LGYVEDIIAAGVADPVKTRBSAVLNASSAVRNVLTFTRVVVDKPAK 524

RESULT 8
Y14909
ID Y14909 standard; protein; 541 AA.

AC Y14909;

DT 25-OCT-1999 (first entry)

DE Amino acid sequence of M. vaccae antigen GV-27.

XX Mycobacterium vaccae protein; antigen; T cell activation; cytokine;
XX dendritic cell maturation; infectious disease; immune disorder; cancer;
XX respiratory system; mycobacterial infection; allergy; tuberculosis;
XX leprosy; sarcoidosis; lung cancer; asthma; skin disorder; psoriasis;
XX dermatitis; eczema; alopecia areata; skin cancer; basal carcinoma;
XX squamous cell carcinoma; melanoma.

OS Mycobacterium vaccae.

PN MO9932634-AA.

XX 01-JUL-1999.

PF 23-DEC-1998; 98WO-NZ00189.
XX
PR 04-DEC-1998; 98US-0205426.
PR 23-DEC-1997; 97US-0996624.
PR 23-DEC-1997; 97US-0997080.
PR 23-DEC-1997; 97US-0997362.
PR 11-JUN-1998; 98US-0095855.
PR 17-SEP-1998; 98US-0156181.
XX
XX (GENE-) GENESIS RES & DEV CORP LTD.
XX Prestige RL, Skinner MA, Tan P, Visser ES, Watson J;
XX WPI: 1999-430163/36.
XX N-PSDB: 211371.
XX
XX Enhancing immune response to an antigen
XX
XX Claim 1; Page 213-214; 243pp; English.
XX
XX The invention provides heat-killed Mycobacterium vaccae, or recombinant
XX M. vaccae proteins. The M. vaccae proteins may be employed to activate
XX T cells and natural killer cells, to stimulate the production of
XX cytokines, to enhance the expression of co-stimulatory molecules on
XX dendritic cells and monocytes, and to enhance dendritic cell maturation
XX and function. The proteins can be expressed by standard recombinant
XX methodology. Pharmaceutical compositions comprising the proteins or
XX nucleic acid sequences encoding the proteins can be used for the
XX treatment, prevention, and detection of disorders including infectious
XX diseases, immune disorders and cancer. In particular, the compounds and
XX methods are used for treatment of diseases of the respiratory system,
XX such as mycobacterial infections, asthma, allergies, tuberculosis,
XX leprosy, sarcoidosis and lung cancers, and disorders of the skin such as
XX psoriasis, atopic dermatitis, eczema, allergic contact dermatitis,
XX alopecia areata, and skin cancers such as basal carcinoma, squamous cell
XX carcinoma and melanoma.

Sequence 541 AA;

Query Match 59.3%; Score 1574.5; DB 20; Length 541;
Best Local Similarity 61.3%; Pred. No. 6.5e-101;
Matches 323; Conservative 79; Mismatches 122; Indels 3; Gaps 2;

1 SKLEIYDFTARAMEVGDKLDVRYTLGPRGRHVLAFAFGPTVNDGTVARETEL 60
2 AKLIAYDEAATRGIERGNAIADAVKVLIPKGRNVLEKKWGAPELNDGVSIAKEI 61
61 EDPEDLGAQLVKSVAATKNDVAGDGTATILAOALIKGLRLVAAGVNFALGVGIGK 120
62 EDPYKIGAEIVKEVAKKTDVAGDGTATLVLGALVREGLRVAAGANPLIGIKYIEK 121
121 AADAVFELASATPVSKGTIAOVATVSSRDEQIGDLVGEAMKVGHDVSVSKESSTL 180
122 AVEKVTCELIGAKVEVEKQIAATAISAGDSIGDLIAEAMDKVNGEYITVEESNLF 181
181 GTELEFTEGIGFHKGFLSAYFVTDPDNOQAVLEDAILILHODKISSLPDLPLEKXVAGT 240
182 GIGLEIPEGMTFIDKYISGYITVDPERGEAVLEDPYILIVSKSVTVKDLIPLEKVI 241
241 GKPLIIAEDVEGALATLVNNAIRKTLKAVAVGKPYGDRRKAFLEDLAVVTGHHVNP 300
242 GKPLIIAEDVEGALATLVNNAIRKTLKAVAVGKPYGDRRKAFLEDLAVVTGHHVNP 301
301 DAGIVLRVGEIVGSAARVYVSKDDIYVDDGGTAEEVANRANHLAEIDKSNDMDRE 360
302 EGVLTLEADISLIGKAKVVKVETDEIVGEGDQDAIGRAVQIQLIENSADYDRE 361
361 KLGRIELAKGAVIVIKGATDPTALKERKESVEDAAVAAKAEVIGPGASLIHQ 420
362 KIGERLAKGAVIVIKGATDPTALKERKESVEDAAVAAKAEVIGPGASLIHQ 420
421 RKALTELASLTGDEVLCVDFSEALAPLFWIANAGLDGSSVVNVKSELPAHGILNVN 480

Db 421 apaldl--gltgdeatganlvrvalsaplkqiafngllepgyvaeksnlpaghnaa 478
 QY 481 TLTSGDLAAGVIDPVKVTSAVLNASSVARMVLTETVVDKRAKA 527
 Db 479 tgyedllkagvadpvkvtksalqnaaslaiflitaavvadkpeka 525

RESULT 9

DB0215
 ID P80215 standard; protein; 560 AA.

XX P80215;

XX 19-MAR-1991 (first entry)

XX Sequence of Mycobacterium tuberculosis 65kD protein.

XX Antigen; vaccine; ds.

XX Mycobacterium tuberculosis.

XX W0805823-A.

XX 11-AUG-1988.

XX 01-FEB-1988; 88WO-US00281.

XX 02-FEB-1987; 87US-0010007.

XX (WHIT-) WHITEHEAD INST BIOM.

XX Husson RN, Young RA, Shinnick TM;

XX WPI; 1988-235175/33.

XX P-PSDB; N80222.

XX Genes encoding Mycobacterium tuberculosis protein antigens -

XX useful for developing reagents for diagnosis, prevention and

XX treatment of tuberculosis

XX Claim 12; Fig 6; 82pp; English.

XX The gene was isolated by probing a lambda gt11 expression library of

XX M.tuberculosis DNA with monoclonal antibodies directed against

XX M.tuberculosis-specific antigens. 19kD, the 71kD and the 65kD proteins

XX and genes are claimed, and so is a vaccine comprising DNA encoding

XX M.tuberculosis protein in a recombinant vaccine vector.

XX Sequence 560 AA;

XX Query Match 59.3%; Score 1574.5; DB 9; Length 560;

XX Best Local Similarity 61.2%; Pred. No. 6.8e-101;

XX Matches 322; Conservative 75; Mismatches 126; Indels 3; Gaps 2;

QY 1 SKLIEYDETRRRAMEVGMKLDATVRYTLCPRGRHNVLAAGFPTVNDGVYARIEL 60

Db 22 aktlaydeartrlgjrlgnladavkvlqpkgrnvlekkwgarpltnvgslakelel 81

QY 61 EDDPEDAGALVSVAKTKTDVAGDGTTPATLLAQAALIKGRLVAGVNPVALGVIGK 120

Db 82 edpyekigaelvkvakttdvagdggtatvagaivregltnvaagapnlgkrglek 141

QY 121 ADAVEFALLASATPVSGKTGAQVATVSSRDEQIGDLVGEAMNKVGHDSVYSKESSTL 180

Db 142 avekvteclllkgakevtekgiaataaisagdsigdlieamdkvgnegvitveesntf 201

QY 181 GTELEFEGEGFHKGLISAFVTFDQNOAVLEDALELLHODKISSLPDLLPLEKXVAGT 240

Db 202 glqleltemrfdkgyisgyfvdpergeavlecdpylllvsskvtstkdlrlplekiga 261

QY 241 GKRLLIIVAEVDEGEALATLVVNAIRKTLKAVAVKGPFGDRRAFLIEDLAVVVGHVNP 300

Db 262 gkpllllaedvegealstlvnkrlygfksavavkapfgdrrkamqlqdmallggqvise 321

QY 301 DAGVLEBVEGLEVLGSRRRVVSXKDDTVIYDGGGTAAVAVNRANHLRAEIDKSDSMDRE 360

Db 322 evgltenadslslgkarkkvvtkdeltivegagddatagrvagrlqgelensdsdydre 381

QY 361 KLGRLAKLAGVAVIRKVGATDTALKERKESVDDAFAAKAAVEEGIVGGASLIHOA 420

Db 382 klgerlaklagvavlaagalevelkerkhrledavrnakaavegivaagvcll-qa 440

QY 421 RKATTELASLTGDEVGVFSEALAPLFMTAANAGLDGSVYVNVKSELPGHGLNVN 480

Db 441 aptldekl--legdeatganlvkaleaplkqiafngllepgyvaeksnlpaghnag 498

QY 481 TLTSGDLAAGVIDPVKVTSAVLNASSVARMVLTETVVDKRAKA 526

Db 499 tgyedlllaagvadpvkvtksalqnaaslaigflitaavvadkpek 544

RESULT 10

W32100
 ID W32100 standard; protein; 540 AA.

XX W32100;

XX 08-APR-1998 (first entry)

XX Mycobacteria sp. heat shock protein 65 (hsp65) #2.

XX Heat shock protein; hsp58; human; autoimmune disease; hsp65;

XX rheumatoid arthritis; antigen; infectious disease; prophylactic;

XX pristanic induced arthritis; PLA; vaccine.

XX Mycobacteria sp.

XX W09711966-A1.

XX 03-APR-1997.

XX 26-SEP-1996; 96WO-GB02382.

XX 27-SEP-1995; 95GB-0019737.

XX (PEPT-) PEPTIDE THERAPEUTIC LTD.

XX Elson CJ, Thompson JS;

XX WPI; 1997-212851/19.

XX Polypeptide(s) derived from microbial heat shock protein - useful

XX for treatment of autoimmune disease esp. arthritis

XX Disclosure; Fig 4; 91pp; English.

XX This is the heat shock protein hsp65 which can be used in a novel method

XX to treat autoimmune disease e.g. rheumatoid arthritis. This sequence is

XX known to be an immunodominant antigen in a number of infectious diseases

XX and is linked to pristanic induced arthritis (PIA) in vitro. HSP's from

XX microbial sources may act as self antigens and thus have limited use

XX whereas the human hsp65 homologue, hsp58 or fragments of the hsp58

XX protein may be useful in the development of vaccines for prophylaxis or

XX treatment of an autoimmune disease such as rheumatoid arthritis.

XX Note: this hsp65 sequence is represented in Figure 4 but differs slightly

XX from the hsp65 sequence given in the sequence specification (see W32099).

XX Sequence 540 AA;

Query Match

Best Local Similarity 59.3%; Score 1572.5; DB 18; Length 540;
 Matches 321; Conservative 76; Mismatches 126; Indels 3; Gaps 2;

QY 1 SKIEIYDEFTARRAMVGVMDKTLADPVRVTYLGRGRHVVLAKAFGGPTTNGCVTARIEL 60

Dd 2 aktlaydeeaatrrglertlinaladavkvtllprkgtnvvlekkwgarblctldngvslakteel 61

OY 61 EDPFEDLCAQLVSVAVTGNTNDVAGDGTATATLLAQAALIKGLRLVAVGVNVALGVIGK 120

Dd 62 eodpykisaenlvkevaktddvagdggttelcvlaegalvreglnvaagapnlpjkgriek 121

OY 121 AADAVEALLASAPRVSGKTGIQAOVAYVSSRDEQIDGLVGEANMKKHDSVYSKESSTL 180

Dd 122 avekvteclikgakvevelkegiaataaisagdqsigdliaeamdkvgnegvltweesntf 181

OY 181 GTELEFETGIGPFHKGFSAFYEVTFPDNOQAVLEDAALILLHQDKTISSLPDLPLEKVAGT 240

Dd 182 gtlletlegmtfdkyuisgytvltdpergeavldepbyllvsaksvtkvdilpllekviya 241

OY 241 GKPLTIIVEDVEGEELATLVVNALTRKTLAAVAAGVPFGDRRKAFLEDLAVTGGHVVNP 300

Dd 242 gkplliiedetvegaelstlivvnkirgltkysavnavparfgdtirkmldmaillbggyise 301

OY 301 DAGIVREVGLEVLGSARRVYSKDITVIYDGGGTAEAVANRANHRAEIDKSDDSDMRE 360

Dd 302 evgtlienadslilgkarkkvutktedrtelliveagdtdaiagrvaqirgelensdsdydre 361

OY 361 KLGEMLATLAGVANITKGATDTDALKRKRESVEDYAAAMAANEGETIVEGGASLIHQ 420

Dd 362 klgerlaklagvavikaagaalevelkerkhrlעדavrnaakaeeiglvaggyvlll-qa 420

OY 421 KRALTELKASTLGDEVLDGVDFSESLAAPLPWIAANAGLDGSSVNNKSELPAHGILNVN 480

Dd 421 aptidalk--legedaatgaanlvkvalaeplrqlqiatfnsglpepgyvaeekvrnlpaghyna 478

OY 481 TLSIGDLAADGVIDPVKYTRSALVNASSAVARMVLTTETVVWDPRAK 526

Dd 479 lgvvydllaagyadvdpvkvtalsalquaasialflitteaevadkpek 524

Db

RESULT 11

W60144

ID W60144 standard; Protein; 523 AA.

AC W60144;

XX

DT 25-AUG-1998 (first entry)

XX

DE M. vaccae antigen GV-27 sequence.

KX

Mycobacterium vaccae; antigen; therapy: prevention; cytokine production;

XX

M. avium; M. tuberculosis; immune response enhancer; cell proliferation;

XX

mycobacteria infection; vaccine; cancer.

OS

Mycobacterium vaccae.

PM WO9806542-A2.

XX

05-MAR-1998.

PD

28-AUG-1997; 97WO-NZ00105.

FE

12-JUN-1987; 97US-0873970.

PR 29-AUG-1996; 96US-0705347.

XX

(GENE-) GENESIS RES & DEV CORP.

PA

Hiyama J, Prestidge RL, Scott LM, Skinner MA, Tan P;

PI Visser E;

XX

WPI; 1998-216926/19.

DR N-PSDB; V34608.

XX

Mycobacterium vaccae polypeptides - used to develop products for use

PT in detection, therapy and prevention of mycobacteria infections or

PT as immune response enhancers

XX Claim 47; Pages 115-116; 153pp. English.

XX This represents a *Mycobacterium vaccae* antigen GV-27. The invention

CC provides M. vaccae polypeptides that comprise an immunogenic portion of

CC a soluble M. vaccae antigen, or a variant, where the antigen induces an

CC immune response in patients previously exposed to a mycobacterium. Such

CC M. vaccae polypeptides can be used in methods for enhancing non-specific

CC immune response. The methods and products can be used for the detection,

CC treatment and prevention of infectious diseases caused by mycobacteria

CC such as M. vaccae, M. avium or M. tuberculosis. The products also have

CC the ability to induce cell proliferation and cytokine production (e.g.,

CC interferon-gamma and interleukin-12 production) in T cells, NK cells,

CC B cells, or macrophages. They can be used for enhancing immune responses

CC for use in vaccines or immunotherapy of infectious diseases and cancers.

XX

XX

XX Sequence 523 AA:

Query Match 59.0%; Score 1566.5; DB 19; Length 523;

Best Local Similarity 61.3%; Pred. No. 2-2e-100;

Matches 321; Conservative 79; Mismatches 121; Indels 3; Gaps 2

QY 1 SKLIEEDERARAMEVGMKLDATYRVTLSPGGRHVLAKAGCGPVTMDGTVARREIL 60

DB :||| ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

2 aktlaydeaartrlgelnaladavkvllpgrygnvlekkwgaprlitldgyslajel 61

QY 61 EDPFEDLGAQLKSVATKNDVADGGTTATTLAQLIKGLRLVAGVNPALGVGIEK 120

DB ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

62 edpyekigselkvkrakktddvaggcttatvtaqalvreglrnvaaagnpiglikrlek 121

QY 121 AADAVFEALLASATPVSGCTGIAOVATVSSRBOICDVLGEAMKNGHDSVSKESSTL 180

DB ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

122 aveavtqslksakeveteqisatcaaisagdtlqigellaeamdkvyngevitveesnl 181

QY 181 GVELEETEGHKGKGLSAYFTDPDNOQAVLEDAILILHODKISSLPRLPLEKVACT 240

DB ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

182 glqlellegmrfdkylisgylvdaergeavledpyllvsskvsckdlrlplekvlga 241

QY 241 GPRLLIVADVEGCEALATLVVNAIRKTLKAVAVKGPFGDRRAKFLIEDAVVTGHIYNP 300

DB ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

242 gprlllilaeedvegeaaltvlvknigtflsvavkbpqfgdrrkamqdmalltgqvase 301

QY 301 DAGIVLEVGLEYLVSARRVVSKDPTVIVDGGTGAEAVANRANHILRAEIDKSDSDMRE 360

DB ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

302 rgvlslsetadvallgarkrvvtldeettiveegsdadlaigvraqiraetiensdsydre 361

QY 361 KIGERLAKIAGVAVYIKGAADPTLAKKEKSEVEDAVAAAKAVEGCIYPGGASLIHQA 420

DB ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

362 kigertlakltagvavlkagaalevelkephkhlledavrnakaaveegiyagsgvall 420

QY 421 RKALTEFLRASLTGDEVLGDFSEALAPLPMIAANAGLDGGSVVVKNVSELPAGCLLNN 480

DB 421 spaldl--glvgdeatgenlrvalsaplkqatfrnglqlegvvaeksnlpagnglnaa 478

QY 481 TLSYGDLAADVIDPYKTRSAVLNASSVARMVLTTEYTVVDKP 524

DB ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

479 tseyedlllkagvadpvyktrsalqnaaslaalfllteavvadkp 522

RESULT 12

ID Y14891

Y14891 standard; protein; 523 AA.

XX Y14891:

XX 25-OCT-1999 (first entry)

XX Amino acid sequence of M. vaccae antigen GV-27.

XX

XX Mycobacterium vaccae protein; antigen; T cell activation; cytokine;

KM dendritic cell maturation; infectious disease; immune disorder; cancer;

KM respiratory system; mycobacterial infection; allergy; tuberculosis;

KM leprosy; sarcoidosis; lung cancer; asthma; skin disorder; psoriasis;
 KM dermatitis; eczema; alopecia areata; skin cancer; basal carcinoma;
 KM squamous cell carcinoma; melanoma.
 OS
 XX Mycobacterium vaccae.
 PN WO9932634-A2.
 PD 01-JUL-1999.
 XX
 XX 23-DEC-1998; 98WO-NZ00189.
 PF
 XX 04-DEC-1998; 98US-0205426.
 PR 23-DEC-1997; 97US-0996624.
 PR 23-DEC-1997; 97US-0997080.
 PR 23-DEC-1997; 97US-0997362.
 PR 11-JUN-1998; 98US-0095855.
 PR 17-SEP-1998; 98US-0156181.
 XX
 XX (GENE-) GENESIS RES & DEV CORP LTD.
 PI Prestidge RU, Skinner MA, Tan P, Visser ES, Watson J;
 DR WPI: 1999-430163/36.
 DR N-PDB: 211343.
 XX
 PT Enhancing immune response to an antigen
 PS
 XX Example 14; Page 189-190; 243pp; English.
 XX The invention provides heat-killed Mycobacterium vaccae, or recombinant
 CC M. vaccae proteins. The M. vaccae proteins may be employed to activate
 CC T cells and natural killer cells, to stimulate the production of
 CC cytokines, to enhance the expression of co-stimulatory molecules on
 CC dendritic cells and monocytes, and to enhance dendritic cell maturation
 CC and function. The proteins can be expressed by standard recombinant
 CC methodology. Pharmaceutical compositions comprising the proteins or
 CC nucleic acid sequences encoding the proteins can be used for the
 CC treatment, prevention, and detection of disorders including infectious
 CC diseases, immune disorders and cancer. In particular, the compounds and
 CC methods are used for treatment of diseases of the respiratory system,
 CC such as mycobacterial infections, asthma, allergies, tuberculosis,
 CC leprosy, sarcoidosis and lung cancers, and disorders of the skin such as
 CC psoriasis, atopic dermatitis, eczema, allergic contact dermatitis,
 CC alopecia areata, and skin cancers such as basal carcinoma, squamous cell
 CC carcinoma and melanoma.
 CC
 XX Sequence 523 AA;
 SO
 Query Match 59.0%; Score 1566.5; DB 20; Length 523;
 Best Local Similarity 61.3%; Pred. No. 2.2e-100;
 Matches 321; Conservative 79; Mismatches 121; Indels 3; Gaps 2;
 QY 1 SKILEYDETRARRAMEVMDKILADTVRVTLGPRGRHYLAFAFGPTVYNGVYARETEL 60
 DB 2 aktlaydeearrjerginaladavkvtlpgpknvlekkwagppltnqsvstakelel 61
 QY 61 EDPFEDIGAOVLVSVATKTDVAGDGTATTALQALIKGGLRLVAGVNPVALGVIGIK 120
 DB 62 edpkyeigaevlkevaktdvagdgtttatvlagalivreglirvaagangpqlkjrglek 121
 QY 121 AADAVFALLASATPVSGKIGIAVATVSSRDGIDLVGEAMKRVHDSVYSKESSTL 180
 DB 122 aveavtllksakevteqisatlaaisagdtqigeliieamdkvngvylveesntf 181
 QY 181 GTELEFEIGIGFHKGFSLAFVTFDNOQAQVLEDPALILHODKISSPDLPLEKXAGT 240
 DB 182 glqleldegmrfdkyisgyfvcdaerqevaledpylllvsksvtkdlplllekviga 241
 QY 241 GKPLLVIAEDVEGBALATLVVNAIRKTLKAAVAVGKPYFGDRRAFLLEDLAVVTGHHVNP 300
 DB 242 gkpllliaedvegealstlvvnkirtgfkavavkapgfdgdrkamldmailtggyvse 301

QY 301 DAGIVREVGLEVLGSAARRVYVSKDPRVYVGGSTAFAVANRANHLEATDKSDPDRE 360
 DB 302 rvglstietadvslllgdrtkvvtldetlvegsqsdalagvaqlaiealsdsdydre 361
 QY 361 KIGERLAKIAGVAVIVGAATDPALEKERKESVEDAAVAAKAAVEGIVPGGASLIHOA 420
 DB 362 klgerlakklggyavivagaaetevelkerhrhiedavrnakaaveegivaggyall-q 420
 QY 421 RKALTELRASTGDEVLGVDVFSALAPLEFWIAANAGLDGSSVYVKNVSELPAGHGLNV 480
 DB 421 apaldldl--gltgdeatganlvralvsaplkqiafngllepgyvaekvsnlpaghylnaa 478
 QY 481 TLSYGDLAADGVDPVAVTSAVINASSVARMVLTETTVVDKP 524
 DB 479 tgeyedllkagvadpvrkvrtslqnaasialfitleavvadkp 522
 RESULT 13
 Y23930
 ID Y23930 standard; Protein; 545 AA.
 AC Y23930;
 XX
 XX 22-SEP-1999 (first entry)
 DE
 DE Consensus amino acid sequence of a heat shock protein.
 XX
 XX Heat shock protein; Hsp; Immune response; Immunological carrier;
 KM cancer control; tumour; sarcoma; cancer; gene therapy.
 XX
 XX Synthetic.
 OS
 XX WO9935270-A1.
 PN 15-JUL-1999.
 PD
 PF 29-DEC-1998; 98WO-CA01203.
 PR 31-DEC-1997; 97US-0001737.
 PA (STRE-) STRESSGEN BIOTECHNOLOGIES CORP.
 PI Mizzen L, Wisniewski J;
 XX
 XX WPI: 1999-430397/36.
 DR
 PT New nucleic acid encoding heat shock protein-60 from Streptococcus,
 PT useful in vaccines, as carriers for other immunogens, as anticancer
 PT agents and for diagnosis
 PS Disclosure; Fig 10A-E; 176pp; English.
 XX
 CC Y23905-30 represent heat shock proteins (Hsps). The specification
 CC describes Streptococcal Hsps, designated Hsp60. These proteins, their
 CC fragments, variants and fusion proteins, are used to elicit or enhance
 CC an immune response against Streptococcus, and to elicit a similar
 CC response to a target antigen fused to the protein. Unlike other
 CC immunological carriers, Hsp60 proteins are not immunosuppressive so
 CC provide an increased response to any conjugated or fused antigen. Also,
 CC where used for cancer control, they lack the side effects associated
 CC with endotoxins. They can also be used to detect specific antibodies
 CC and in treatment or prevention of tumours (e.g. sarcoma or cancers of
 CC breast, ovary, prostate, lung, pancreas or liver). The Hsp60
 CC polynucleotide is used for recombinant production of the protein, as
 CC a source of primers and probes for detecting streptococci in standard
 CC hybridization/amplification assays, and therapeutically in gene
 CC therapy vectors.
 CC
 XX Sequence 545 AA;
 SO
 Query Match 58.6%; Score 1555.5; DB 20; Length 545;

PS Disclosure; Page 69-71, 91pp; English.
XX

CC This is the heat shock protein hsp65 which can be used in a novel method
CC to treat autoimmune disease e.g. rheumatoid arthritis. This sequence is
CC known to be an immunodominant antigen in a number of infectious diseases
CC and is linked to pristane induced arthritis (PIA) in vitro. HSP's from
CC microbial sources may act as self antigens and thus have limited use
CC whereas the human hsp65 homologue, hsp58 or fragments of the hsp58
CC protein may be useful in the development of vaccines for prophylaxis or
CC treatment of an autoimmune disease such as rheumatoid arthritis.
CC Note: this sequence is given in the sequence specification but differs
CC slightly from the hsp65 sequence represented in Figure 4 (see W32100).
XX
SQ Sequence 544 AA;

Query Match 58.6%; Score 1554.5; DB 18; Length 544;
Best Local Similarity 60.4%; Pred. No. 1.6e-99;
Matches 320; Conservative 76; Mismatches 127; Indels 7; Gaps 3;

```
0Y      1 SKLIYEDTARRAMEVGMKLDADVRLVLPGRHRHVLAKAFCGCTVTNDGVAREIEL    60
       : | | | | : | : | : | | | | | | | | : | | | | | : | | |
Db      2 aktiaydearriglerglnaladeakvblgpkgrnrvlekkwgpbtindgvstakejel    61
       : | | | | : | : | : | | | | | | | | : | | | | | : | | |
0Y      61 EDPEFDLGAOLVKSVATYTNDVAGDCTTTATILAOALIKGLRLVAAGVPVALCVGTGK   120
       ||| : | | | | | | | | | | | | | | | | | | | | | | | |
Db      62 edpyekigaelykveakkrtddvaggtcttaclaqalvreglrmaagapnlgikrglek   121
       ||| : | | | | | | | | | | | | | | | | | | | | | | | |
0Y      121 AADAVEALLIASAPFVSCKTGTAQAVATYSNR --- EDIGLVCEAMKVGHDSVASKE   176
       ||| : | | | | | | | | | | | | | | | | | | | | | | | |
Db      122 avekvtelllygakvetkeqiaalaasagdsigdgljaeamkvngelyllvee     181
       ||| : | | | | | | | | | | | | | | | | | | | | | | | |
0Y      177 SSTLOTLEFEETGEOHKGFLTSAYEVTDFDNOQAAVLDELALLLLHQDKISSPDLPLER    236
       ||| : | | | | | | | | | | | | | | | | | | | | | | | |
Db      182 snltfglqietegmfdkgyslyfvtdpergeavleedpylllvssksstvkdllptlek    241
       ||| : | | | | | | | | | | | | | | | | | | | | | | | |
0Y      237 VAGTGPPLLIAEDVEGALATLVNNAIRKLKAAVANGPYFGDRRKAFLEDLAVVYGGH    296
       | | | | | | | | | | | | | | | | | | | | | | | | | |
```

Dd	242	vlgagkplllllaedvegalsltlvnkkrlgtfkfsavvkapfgdrirkamqlmdalllbggq	301
Oy	297	VYNPAGVILREVGCEVLGSGARRVVNSKDDPIVIYDGGCTAAEAVNRRANHLRAELDKSDS	356
Dd	302	vlseevgltilenadslilgkarvkvvcldectivegagdtaigvaqilqelensod	361
Oy	357	WDREKLGERLAKLAGVAIVIRVGATDTALRKERESVEDAFAAKAAVEEGIPGCGASL	416
Dd	362	ydreklgerlaklagvavikagaatevelkerkhriedavrnakaaveegivaggvtl	421
Oy	417	IHOAKKALTLELRASLTGGEVLGVDFESALAAPLFMTIAANGLDGSSVYNNKVSSELPACHG	476
Dd	422	l-qaaprtldalk--legdeatgaanlvkaalaplkgiafnsglpepyvaaekvrnlpgahg	478
Oy	477	LNVNPLTSGYDAADGVIDPVVTTSAYTSRNASSVARMLVTTEVVVDKRPK	526
Dd	479	lnagtgyvedllaagvadpvkvrtlsalnnaasigilfilitcaavvadpkex	528
RESULT 15			
XX	R04716		
XX	R04716 standard; protein; 572 AA.		
XX	R04716;		
XX	03-JUL-1990 (first entry)		
DE	Amino acid sequence of the 65 kDa Mycobacterium tuberculosis protein.		
XX	Stress protein; heat shock protein; Mycobacterium tuberculosis protein;		
KW	vaccine; human P1 protein; Mycobacterium; immunotherapeutic agent.		
XX	Homo sapiens.		
OS			
XX	MOB912455-A.		

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 12, 2001, 18:01:13 ; Search time 3358.72 Seconds
(without alignments)
3254.716 Million cell updates/sec

Title: US-09-461-774-3

Perfect score: 1560
Sequence: 1 gcgcgcattggtgcgcgcgcag.....gtctgcgtgcgcgcctga 1560

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 7991742 seqs, 3503743858 residues

ctal number of hits satisfying chosen parameters: 15983484

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: em_est4a:*
42: em_estfun:*
43: em_esthum1:*

44: em_esthum2:*
45: em_esthum3:*
46: em_esthum4:*
47: em_esthum5:*
48: em_esthum6:*
49: em_esthum7:*
50: em_esthum8:*
51: em_esthum9:*
52: em_esthum10:*
53: em_esthum11:*
54: em_esthum12:*
55: em_esthum13:*
56: em_esthum14:*
57: em_esthum15:*
58: em_esthum16:*
59: em_esthum17:*
60: em_esthum18:*
61: em_esthum19:*
62: em_esthum20:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estov1:*
68: em_estov2:*
69: em_estp11:*
70: em_estp12:*
71: em_estp13:*
72: em_estp14:*
73: em_estp15:*
74: em_estro1:*
75: em_estro2:*
76: em_estro3:*
77: em_estro4:*
78: em_estro5:*
79: em_estro6:*
80: em_estro7:*
81: em_estro8:*
82: em_estro9:*
83: em_estro10:*
84: em_estro11:*
85: em_estro12:*
86: em_estro13:*
87: gb_est41:*
88: gb_est42:*
89: gb_est43:*
90: gb_est44:*
91: gb_est45:*
92: gb_est46:*
93: gb_est47:*
94: gb_est48:*
95: gb_est49:*
96: gb_est50:*
97: gb_est51:*
98: gb_est52:*
99: gb_est53:*
100: gb_est54:*
101: gb_est55:*
102: gb_est56:*
103: gb_est57:*
104: gb_est57:*
105: gb_est58:*
106: gb_est59:*
107: gb_est70:*
108: gb_est71:*
109: gb_est72:*
110: gb_est73:*
111: gb_est74:*
112: em_esthum21:*
113: em_esthum22:*
114: em_esthum23:*
115: em_estom1:*
116: em_estom2:*

190: gb_gss25:*
191: gb_gss26:*
192: gb_gss27:*
193: gb_gss28:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
117: em_estp16:*						
118: em_estp17:*						
119: em_estp18:*						
120: em_estp14:*						
121: em_estp15:*						
122: em_estp16:*						
123: em_estp17:*						
124: em_estp18:*						
125: em_estp19:*						
126: gb_est58:*						
127: gb_est59:*						
128: gb_est60:*						
129: gb_est61:*						
130: gb_est62:*						
131: gb_est63:*						
132: gb_est64:*						
133: gb_est65:*						
134: gb_est66:*						
135: gb_est67:*						
136: gb_est68:*						
137: gb_est69:*						
138: gb_est70:*						
139: gb_est71:*						
140: gb_est72:*						
141: gb_est73:*						
142: gb_est74:*						
143: gb_est75:*						
144: gb_est76:*						
145: gb_est77:*						
146: gb_est78:*						
147: gb_est79:*						
148: gb_est80:*						
149: gb_est81:*						
150: gb_est82:*						
151: gb_est83:*						
152: gb_est84:*						
153: gb_est85:*						
154: gb_est86:*						
155: gb_est87:*						
156: gb_est88:*						
157: gb_est89:*						
158: gb_est90:*						
159: gb_est91:*						
160: gb_est92:*						
161: gb_est93:*						
162: gb_est94:*						
163: gb_est95:*						
164: gb_est96:*						
165: gb_est97:*						
166: gb_est98:*						
167: gb_est99:*						
168: gb_est100:*						
169: gb_est101:*						
170: gb_est102:*						
171: gb_est103:*						
172: gb_est104:*						
173: gb_est105:*						
174: gb_est106:*						
175: gb_est107:*						
176: gb_est108:*						
177: gb_est109:*						
178: gb_est110:*						
179: gb_est111:*						
180: gb_est112:*						
181: gb_est113:*						
182: gb_est114:*						
183: gb_est115:*						
184: gb_est116:*						
185: gb_est117:*						
186: gb_est118:*						
187: gb_est119:*						
188: gb_est120:*						
189: gb_est121:*						

ALIGNMENTS

RESULT 1
CNS0091P/c
LOCUS
DEFINITION
Drosophila melanogaster genome survey sequence TET3 end of BAC #
fly), genomic survey sequence.
ACCESSION
VERSION
KEYWORDS
GSS.
GSS.

190: gb_gss25:*
191: gb_gss26:*
192: gb_gss27:*
193: gb_gss28:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	64.6	4.1	925	190	CNS0091P	AL053013 Drosophila
2	64.2	4.1	932	190	CNS00720	AL066742 Drosophila
3	63.6	4.1	925	190	CNS0091P	AL053013 Drosophila
4	61.4	3.9	932	190	CNS00720	AL066742 Drosophila
5	58	3.7	839	190	CNS004NB	AL054280 Drosophila
6	57.4	3.7	953	94	AW727549	AW727549 GA_Ea001
7	57.2	3.7	1315	94	AW731134	AW731134 GA_Ea001
8	56.2	3.6	935	190	CNS006XK	AL066051 Drosophila
9	56.2	3.6	1212	94	AW727500	AW727500 GA_Ea001
10	54	3.5	1203	190	CNS015Y4	AL106054 Drosophila
11	54	3.5	1338	94	AW731160	AW731160 GA_Ea001
12	53	3.4	1063	94	AW727530	AW727530 GA_Ea001
13	52.4	3.4	935	190	CNS006XK	AL066051 Drosophila
14	52.4	3.4	1158	94	AW727550	AW727550 GA_Ea001
15	52	3.3	839	190	CNS004NB	AL054280 Drosophila
16	51.8	3.3	1009	190	CNS010EW	AL098882 Drosophila
17	51.6	3.3	1009	190	CNS010EW	AL098882 Drosophila
18	51.4	3.3	761	176	A2184492	A2184492 SP_1003_A
19	51.4	3.3	886	176	A2196424	A2196424 SP_1032_A
20	51.2	3.3	1201	190	CNS01523	AL106089 Drosophila
21	51	3.3	921	176	A2211117	A2211117 SP_015_B
22	51	3.3	1101	190	CNS0175Y	AL108460 Drosophila
23	50.2	3.2	1203	190	CNS015Y4	AL106054 Drosophila
24	49.4	3.2	504	95	AW769372	AW769372 h163b07.x
25	49.4	3.2	955	190	CNS006SE	AL069794 Drosophila
26	49.4	3.2	1010	190	CNS016TE	AL107180 Drosophila
27	49.2	3.2	645	190	CNS01213	AL101589 Drosophila
28	49.2	3.2	910	190	CNS006ON	AL065629 Drosophila
29	49.2	3.2	1101	190	CNS0175Y	AL108460 Drosophila
30	49	3.1	841	176	A2196416	A2196416 SP_1031_B
31	49	3.1	1100	190	CNS016KD	AL106855 Drosophila
32	49	3.1	1347	94	AW731204	AW731204 GA_Ea001
33	48.8	3.1	844	190	CNS0052P	AL056652 Drosophila
34	48.4	3.1	1415	94	AW727479	AW727479 GA_Ea001
35	48.2	3.1	879	191	CNS020OG	AL209545 Tetradon
36	48	3.1	465	136	BE857943	BE857943 7172f02.x
37	48	3.1	539	23	A1652297	A1652297 wb28g05.x
38	48	3.1	870	190	CNS006EZ	AL064271 Drosophila
39	48	3.1	879	191	CNS020OG	AL209545 Tetradon
40	48	3.1	1009	190	CNS017JA	AL108112 Drosophila
41	48	3.1	1622	94	AW731133	AW731133 GA_Ea001
42	47.6	3.1	1099	94	AW731227	AW731227 GA_Ea001
43	47.4	3.0	1236	106	BE319582	BE319582 NF015D07R
44	47.2	3.0	933	94	AW731165	AW731165 GA_Ea001
45	47.2	3.0	936	176	A2186391	A2186391 SP_1006_B

SOURCE	fruit fly.
ORGANISM	Drosophila melanogaster
REFERENCE	Euarystota, Melazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
AUTHORS	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
TITLE	Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
JOURNAL	1 (bases 1 to 925)
COMMENT	Genoscope. Direct Submission Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr) Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammossier at Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPc1-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES	Location/Qualifiers
source	1..925
	/organism="Drosophila melanogaster"
	/db_xref="taxon:7227"
	/clone_lib="RPc1-98"
	/clone="BACR19D16"
	/note="end : TET3"
BASE COUNT	120 a 61 c 61 g 172 t 511 others
ORIGIN	
Query Match	4.1%; Score 64.6; DB 190; Length 925;
Best Local Similarity	14.0%; Pred. No. 7.8e-05;
Matches	50; Conservative 170; Mismatches 137; Indels 0; Gaps 0;
OY	840 ctaccgcccgcagctcgccgcgttcgcggcctcgtgcgcacagctgcgcagtg 899
Db	913 SBSCSSSMSTSNSSBSCSSCBSSBSTSMSSSBSSSGSSGTTSAVKKN 854
OY	900 gtccaacgcctaccgccctgtgtgaacctgtgtgcagaagccggatgaagctgcga 959
Db	853 ASSSGCGCCGCAGBCMCSSSSSCGSASARVKRASGADAKRGCGSGAASHSSSA 794
OY	960 tcacagtgcctgaactaacgcgcagcagcagcaggcaacatcaacgcgctgtacagcc 1019
Db	793 CBSSSSSCSASCWSASSSSSSABSSRSRSGCAGGAGGASSRSSSSASAGSVSSAS 734
OY	1020 tcagcgtcggaagtaccctgcagcacgagtgtctgttgggtgtgcagcgcgcagcagcgcgg 1079
Db	733 SSSSCSSSVSCSVNASMSCSSBSSSSSASASSSSSSASACSCCCTSWGSCSTSA 674
OY	1080 cgactgtgtgtgtcgttcgcagcagactabaagcgcgggatgcagcggcacttaagcaa 1139
Db	673 MSAAARSSSSSSSSSSMSASSSASSSSSSSSGSSSSSGSACGBMSSGGSGSVS 614
OY	1140 cgaccagcagcgttacaagcgttcggtcgtctcatgtgcgccacacgcgcagatcc 1196
Db	613 ASSGSSSVSSSGGSGSGGGGAGVGSGSSSSSGSGSGSGSVCSGCCMCRCSC 557
RESULT 2	CNS00720 932 bp DNA GSS 03-JUN-1999
LOCUS	Drosophila melanogaster genome survey sequence T7 end of BAC #
DEFINITION	BACR14B09 of RPc1-98 library from Drosophila melanogaster (fruit
ACCSSION	fly), genomic survey sequence.
	AL066742

VERSION	AL06742.1	GI:4945205
KEYWORDS	GSS.	
SOURCE	fruit fly,	
ORGANISM	Drosophila melanogaster Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila. 1 (bases 1 to 932)	
REFERENCE	Genoscope.	
AUTHORS	Direct Submission	
TITLE	Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage	
JOURNAL	Bp 191 91006 EVRY cedex - FRANCE (E-mail : seqrefgenoscope.cns.fr	
COMMENT	- Web : www.genoscope.cns.fr) Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammeter in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPc1-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2: cn bw sp, the same strain used for the BDGP's pl and Est libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.	
FEATURES	location/Qualifiers	
source	1..932 /organism="Drosophila melanogaster" /db_xref="taxon:7227" /clone_lib="RPc1-98" /clone="BACR1AB09" /note="end : 77"	
BASE COUNT	155 a 202 c 241 g 91 t 243 others	
ORIGIN		
Query Match	4.1%; Score 64.2; DB 190; Length 932;	
Best Local Similarity	33.0%; Pred. No. 9,6e+05;	
Matches 151:	Conservative 92; Mismatches 211; Indels 3; Gaps 1;	
OY 1014	cagccctcagcgctggaagtacttaccagtagtgctgtctgggagtcgcagcgcgagcaga 1073	
Db 469	CKKCBGCGKGHBKGGAGGKBGCKGCGGCGGGGGVGCGRGGSGSGCSTMGCGGCK 528	
OY 1074	cgcgcgacctgtgtgtgtgtcttgcgcgaactatgacggcgcgatctgcagacygcacta 1133	
Db 529	GGCGCGCTSKMCGTGBGSTGCGGGGGSXSGSGSGCSBHCGGGSGTGMGSGGGGSPBSB 588	
OY 1134	cagcaagcagcgagcgagcgattcaaacgagtcgcggtccggtgtcatatgcccacaacaggcca 1193	
Db 589	TKTYBKRGKGGCGGSSGCCSGSCSGKGGGGCCSCGSGGCGGGSSCGCCSCGCGSG 648	
OY 1194	tccagcgagctgtgtgtgtgcgcgcgcacaaagatccgtcaggtcgcgcccttactagta 1253	
Db 649	GCGCGCGCGSSSGSGSGSGSGSGSGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 708	
OY 1254	cgggcagttcaccgatccgcgcc--cgcatctgtgcgcgctgtgacgttgcgcgttcgcgc 1310	
Db 709	CGGSGCGCGCGSGCGSCGCCSCGCCSCGCCSCGCCSCGCCSCGCCSGSGSGSGSG 768	
OY 1311	aacgtcagcgccgacccccgc 1370	
Db 769	CGCGCGGSGCGSGCGCGCGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSG 828	
OY 1371	caccacacacccgcgcacatcgtatcagtcgcgggtagaccgcgcgcgcgcgcgcgcgcgc 1430	
Db 829	SGCGSGSGGCG 888	
OY 1431	accgctgatcaccttcgacgcgcgcacacacacatcgcgcgcgcgcgcgcgcgcgcgcgc 1467	

Db	889	SCGGCGGSCGSCGSCCGSSSSSGCCGCGGGGG	925
RESULT	3		
LOCUS	CNS0091P		
DEFINITION	CNS0091P 925 bp DNA Drosophila melanogaster genome survey sequence TERT end of BAC # BACR1D16 of Rpci-98 library from Drosophila melanogaster (fruit fly) genomic survey sequence.	GSS	03-JUN-1999
ACCESSION	AL053013		
VERSION	AL053013.1	GI:4934461	
KEYWORDS	GSS.		
SOURCE	fruit fly.		
ORGANISM	Drosophila melanogaster		
REFERENCE	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.		
AUTHORS	1 (bases 1 to 925)		
TITLE	Genoscope.		
JOURNAL	Direct Submission		
COMMENT	Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - web : www.genoscope.cns.fr)		
	Determination of this BAC-end sequence was carried out as part of a		

and BACs for constructing a physical map of the *Drosophila* melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP *Drosophila* melanogaster BAC library was prepared by Kazutoko Osoegawa and Aaron Mammosser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPc1-98 and was constructed by partial EcoRI digestion of *Drosophila* DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

Location/Qualifiers

	BASE COUNT	ORIGIN
/organism="Drosophila melanogaster"	120 a	61 g
/db_xref="taxon:7227"		172 t
/clone_lib="RPCI-98"		511 others
/clone="BACR19D16"		
/note="end : TET3"		

Query Match	4.1%;	Score 63.6;	DB 190;	Length 925;
Best Local Similarity	12.7%;	Pred. No. 0.00013;		
Matches 46;	Conservative 177;	Mismatches 139;	Indels 0;	Gaps 0;
QY 1074	cgccgcgcagctgtcgtgtgtcgtccgcagactatcagcgccggatgacagagcacta	1133		
Db 555	SSSGYKYGKCGSSGSBSBSCCSCSSSCCCGCCCCCCGCCSSSVCCSSBSBSCSSTS	614		
QY 1134	cagcaacaccaggaagcgcgttcacaagcgcgtgcgtgtcgtatcgtccacacgcgcga	1193		
Db 615	BSCCSCGSKYCVCGTSCSSSSSSSSSSSTSSSTSSSTSSKSSSGSSSSSVTTKMS	674		
QY 1194	tccagcgccctcgggtgtgcgcgcgcgaacacagatcctcgtagtgcgcccttcattagcta	1253		
Db 675	TSASGSGWSMAGGSGSTGTSTSSSSSSSTSTSSSSSVSGSKSTBSBSSGSSSSSSS	734		
QY 1254	cgggcagttcacccgatactccgcccccccgagatctgtgctgcgtctgtgcccgtgcccgaac	1313		
Db 735	STSSBBSCSTSTSSSSSSSVSSSTSCCTCCCSYSTSSSTSSSSSTSMGSSSSVGT	794		
QY 1314	gtcagccagccagccagccggcgccgcgcgcggggatctgcgaaggtcgtggtgtccacac	1373		
Db 795	SSSSDSTSTCCSCCCCTACCTCTBMBCTYTSTSCGSSSSSGKGGVTKCGGGCGSSSTN	854		

Qy	1374	cgaagcaggccactctcgatcatgiccgagggtagacccgcggccgcgcagcttgggcacc	1433
		: ::: ::: ::: ::: ::: ::: ::: ::: :::	
Dd	865	GMBTSSACSSSSSCSSSVSSSKRSAASSSVSSSGSVSSNSSSAAKSSSSGSIVSS	914
Qy	1434	gc	1435
		::	
Dd	915	GS	916

RESULT	4			
CNSD0720/c				
LOCUS				
DEFINITION	CNSD0720	932 bp	DNA	GSS 03-JUN-1999
	Drosophila melanogaster genome survey sequence T7 end of BAC #			
	BACR1AB09 of RpCI-98 library from Drosophila melanogaster (fruit			
	fly), genomic survey sequence.			
ACCESSION	AL066742			
KEYWORDS	AL066742.1	GI:4945205		
SOURCE	GSS.			
	Fruit fly.			

REFERENCE
AUTHORS
TITLE
JOURNAL

1 (bases 1 to 932)
Genoscope.
Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage

COMMENT

Determination of this BAC-end sequence was carried out as part of collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the *Drosophila melanogaster* genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP *Drosophila melanogaster* BAC library was prepared by Kazuhiro Oseegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPc1-98 and was constructed by partial EcoRI digestion of *Drosophila* DNA provided by the BDGP from the isogenic strain Y2: cn bw sp, the same strain used for the BDGP's p1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm

```

source
1. 932
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_1id="RPCI-98"
/clone="BACR14B09"
/note="end : T7"
BASE COUNT 155 a 202 c 241 g 91 t 243 others
ORIGIN

```

[illegible]

[illegible]

RESULT 7
AW731134
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Gossypium arboreum.
Gossypium arboreum
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II
Malvales; Malvaceae; Gossypium.
1 (bases 1 to 1315)
Leslie,A., Frisch,D., Yu,Y., Wood,T.C., Wing,R.A. and Wilkins,T.A.
An integrated analysis of the genetics, development, and evolution
of the cotton fiber
Unpublished (2000)
JOURNAL
COMMENT
Contact: Wing RA
Clemson University Genomics Institute
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
High quality sequence stop: 1315.
Location/Qualifiers
1..1315

```

/organism="Gossypium arboreum"
/strain="AKA"
/cultivar="8400"
/db_xref="taxon:29729"
/clone_ga="Ea0010A03"
/clone_lib="Gossypium arboreum 7-10 dpa fiber library"
/tissue_type="Fibers isolated from bolls harvested 7-10 dpa"
/lab_host="E. coli"
/ncbi_vector="pK-CMV; Site_1: EcoRI; Site_2: XhoI"

```

BASE COUNT	78 a	777 c	348 g	101 t	11 others
ORIGIN					

Query Match	3.7%	Score 57.2;	DB 94;	Length 1315;
Best Local Similarity	44.3%	Pred. No. 0.0038;		
Matches 274; Conservative	0;	Mismatches 344;	Indels 1;	Gaps 1;

[illegible]

RESULT 8
CNS006KX/c
LOCUS CNS006KX 935 bp DNA GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC #
BACR14N09 of RPC1-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL066051
VERSION AL066051.1 GI:4945019
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 935)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage ;
BP 121 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)

KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Plasmid Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachyera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1203)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
http://www.edgp.ebi.ac.uk . This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billaud at CEPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
peloBAC11.
FEATURES
source location/Qualifiers
1..1203
/organism="Drosophila melanogaster"
/plasmid="pBelobAC11"
/db_xref="taxon:7227"
/clone_lib="DrosBAC"
/clone="BACN15E10"
/note="end : 17"
BASE COUNT 154 a 274 c 380 g 158 t 237 others
ORIGIN
Query Match 3.5%; Score 54; DB 190; Length 1203;
Best Local Similarity 32.7%; Pred. No. 0.02;
Matches 156; Conservative 77; Mismatches 244; Indels 0; Gaps 0;
QY 5 ccattgtgagccgacagacgcttgagctgcgctgtcttccttcggtgtgtctgcgcg 64
DB 701 CCSSGG 760
QY 65 gactgtccttagcagcgcccccgcttgccgctgcgactgaagaccgagcgccgcaaa 124
DB 761 CC 820
QY 125 ccccggtgagcgcgctgtgtgcgcgcgcaacagcttgaaacgtgcgcgagttacgt 184
DB 821 CGGSSGGGGGCGSSGG 880
QY 185 ccgaacacagcgaatcgcacgtgcgcgacgtgtgcgttcctcccgctgactacg 244
DB 881 GGGGGGSSCGCGSSGG 940
QY 245 accaaccggttggaacacagcgaagtgtgcgtgtatccgcgttcgcgcgagcgagc 304
DB 941 GGSAGGGMVGMAGGGGGGSCCMGIVSSSVSGSSGSGGGGGGGGGGGGGGGGGGVM 1000
QY 305 gattcgaagacgtgtatcattccttgaggagaccgagcgctgcgcgtgcacatgtcg 364
DB 1001 GGGGGGGGGCGSSSSGGGGGMSGMSNMGGGGGMSGSSSSSSGGGGGGGGGGGGCC 1060
QY 365 ccgcatgacacgcgcgcatgcgcgacacgacattctcgcgcaattcgcactgtgtgct 424
DB 1061 CSSSSSSGSCSCSCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCC 1120
QY 425 tcgacccgaagaggtgcgcgactgcacccctcgtgtgcgtgtgcacccgaagcg 481
DB 1121 GCGCGCGGGGGCGCCCGCGCGCCCGCCCGCGSGGSGSCSGSGGGGCS 1177

ACCESSION arboresum cdna clone GA_Ea0010E07, mRNA sequence.
VERSION AW731160
AW731160.1 GI:7628818
KEYWORDS EST.
SOURCE Gossypium arboreum.
ORGANISM Gossypium arboreum
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
Malvales; Malvaceae; Gossypium.
REFERENCE 1 (bases 1 to 1338)
AUTHORS Leslie, A., Frisch, D., Yu, Y., Wood, T. C., Wing, R. A. and Wilkins, T. A.
TITLE An integrated analysis of the genetics, development, and evolution
of the cotton fiber
JOURNAL Unpublished (2000)
COMMENT Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
High quality sequence stop: 1338.
FEATURES
source location/Qualifiers
1..1338
/organism="Gossypium arboreum"
/strain="AKA"
/cultivar="8400"
/db_xref="taxon:29729"
/clone_lib="GA_Ea0010E07"
/clone="GA_Ea0010E07"
/tissue_type="Fibers isolated from bolls harvested 7-10
dpa"
/lab_host="E. coli"
/note="Vector: PBK-CMV, Site_1: EcoRI, Site_2: XhoI"
BASE COUNT 46 a 597 c 619 g 8 t 8 others
ORIGIN
Query Match 3.5%; Score 54; DB 94; Length 1338;
Best Local Similarity 44.7%; Pred. No. 0.02;
Matches 210; Conservative 0; Mismatches 260; Indels 0; Gaps 0;
QY 888 ctcgcccagttggtcaacgactacacgcccctgtgtgacgctgtgtgagaagccg 947
DB 602 CGCGGGCG 661
QY 948 taagaagctggatcagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1007
DB 662 CG 721
QY 1008 gctgtacagccctcagcgtgcgtggaagtacacctgacacgtgtgtgtgtgtgtgtgt 1067
DB 722 CCG 781
QY 1068 cagcagcgcgcgcgactgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1127
DB 782 CCGCGGGGGCG 841
QY 1128 gacttaagaaagacagcagcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1187
DB 842 CG 901
QY 1188 ggcgcgtacagcgcgcgt 1247
DB 902 GCGACG 961
QY 1248 tagctacgagcagttacacgatacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1307
DB 962 CGCGGGCG 1021
QY 1308 ggcacagctgacgc 1357
DB 1022 CG 1071

QY	68	tgctctctaaagcagcccccgcttgcgcgagcgactgaagaacccgagccgcgccaacac	127
	582	5BBSVSTGCTCTCTKCSGCCSTGSGCTGCCCCGGGGSGCCGCGCGCCSGGGCGGCCS	641
QY	128	cgagtgccgcgcggtlctgtbgtccgcacaaagatttgaaacagactgtccgcgagttcatcgcg	187
Db	642	SGSGSGSSGSGSGSGSSSGS	701
QY	188	aaaccaagcgaaattgcgactgcagctgcgcgagcggltgcgttcccgctgcactagacc	247
Db	702	GGSGSGCGSGSGGCGSGCGSCGSCGSGCGSGSGSGSGCGSCCGSCGSCSSSGCGCGCGC	761
QY	248	aaccgcgttgtagacacaaaggaagtgtgcgtgtatccgcgttcccgcgagacagcagat	307
Db	762	GMSKSGCGCGSGSGCG	821
QY	308	tcgagacgactgtgtlcaatctctgtggggagccggggggtgcgtgcggtlccgacatgtgcg	367
	822	SGSGCGSGCGCGCGCGCGS	881
QY	368	ctatggacaccgcgagctgcgcgacacgcgaattctcgcgcaactgacc	415
Db	882	SGSGSGSGCGSGCC	929
RESULT 14			
LOCUS	AM727550/c		
DEFINITION	GA_AM727550 1158 bp mRNA EST 20-APR-2000		
ACCESSION	GA_Ea0012008 Gossypium arboreum 7-10 dpa fiber library Gossypium		
VERSION	AM727550		
KEYWORDS	AM727550.1 GI:7625106		
SOURCE	Gossypium arboreum.		
ORGANISM	Gossypium arboreum		
REFERENCE	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eucosids II; Malvales; Malvaceae; Gossypium.		
AUTHORS	1 (bases 1 to 1158)		
TITLE	Leslie,A., Fritsch,D., Yu,Y., Wood,T.C., Wing,R.A. and Wilkins,T.A.		
JOURNAL	An integrated analysis of the genetics, development, and evolution		
COMMENT	of the cotton fiber unpublished (2000)		
	Contact: Wing RA		
	Clemson University Genomics Institute		
	Clemson University		
	100 Jordan Hall, Clemson, SC 29634, USA		
	Tel: 864 656 7288		
	Fax: 864 656 4293		
	Email: rwing@clemson.edu		
FEATURES	High quality sequence stop: 1158.		
SOURCE	Location/Qualifiers		
	1..1158		
	/organism="Gossypium arboreum"		
	/strain="AKA"		
	/cultivar="8400"		
	/db_xref="taxon:29729"		
	/clone="GA_Ea0012008"		
	/clone_lib="Gossypium arboreum 7-10 dpa fiber library"		
	/issue_type="Fibers isolated from bolls harvested 7-10		
	dpa"		
	/lab_host="E. coli"		
	/note="vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"		
BASE COUNT	70 a 530 c 485 g 64 t 9 others		
ORIGIN			
Query Match	3.4%; Score 52.4; DB 94; Length 1158;		
Best Local Similarity	43.1%; Pred. No. 0.047;		
Matches	245; Conservative 0; Mismatches 323; Indels 0; Gaps 0;		
17	gcagacgltgagctcgcgctgtctgtlctccttcggtlctgctgcgcgagctctctag	76	

[illegible]

filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES

Source

1. 839
Location/Qualifiers

/organism="Drosophila melanogaster"

/db_xref="taxon:7227"

/clone_lib="RPC1-98"

/clone="BACR10E16"

/note="end : TMT3"

BASE COUNT 285 a 67 c 77 g 27 t 383 others

ORIGIN

Query Match

3.3%; Score 52; DB 190; Length 839;

Best Local Similarity 16.5%; Pred. No. 0.057; Mismatches 146; Indels 0; Gaps 0;

Matches 58; Conservative 148; Mismatches 146; Indels 0; Gaps 0;

```
QY 1016 gccctcagcgctggaagctacgtacagctgctgctgctgagcgctgacgagcgagcgagc
DB 838 ssobbrtgcgssrtbbsccsctssssstsssktgcgcsksksbsbsgssssscgcttysc
QY 1076 ccgagcacttgctgctgctgctgacgacactatgacggccggatgcagacggcactaca
DB 778 stgsgsgssvsgsgsgsgsgsgsgsgsgsgsgsgsgsgsgsgsgsgsgsgsgsgsgsgsg
QY 1136 gcaacgacgacgacgacgcttcaacgcgctcgctatgctgccccacacgagcgagc
DB 718 tscgsssttcsbtrgksssgsgsgsgsgsgsgsgsgsgsgsgsgsgsgsgsgsgsgsgsg
QY 1196 cagcgagcctggtgagcgcgacgacgagatccgctcaggtcgccccgttccctagctacg
DB 658 tttststtsttsttsttsccttckckcvtcvtbtttsbcbcsccsgbttbscsss
QY 1256 ggcagttcaacggagatccgcccccgacgtgtgagcgctgtgagccggtgcgagcaagtc
DB 598 ssbgsttsssssssgskgcsstssststsgsgsgsgsgsgsgsgsgsgsgsgsgsgsgsg
QY 1316 cgaacgacgacgacgacgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagc
DB 538 cscgtccscyscstccscctcgsgcsttssscsttscstgtsctgttbsggttb 487
```

Search completed: January 12, 2001, 18:01:32
Job time: 21228 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 12, 2001, 20:09:43 ; Search time 153.29 Seconds
(without alignments)
1640.094 Million cell updates/sec

Title: US-09-461-774-3

Perfect score: 1560

Sequence: 1 ggcgcacatgagcgcgcag.....gtctgagtgagcgcctga 1560

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 280836 seqs, 80580151 residues

tal number of hits satisfying chosen parameters: 561672

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6.COMB.seq:*
4: /cgn2_6/ptodata/1/ina/PCIOS.COMB.seq:*
5: /cgn2_6/ptodata/1/ina/Backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	265.2	17.0	1820	1	US-08-173-508-7
2	265.2	17.0	1821	2	US-08-265-310-7
3	265.2	17.0	1821	3	US-08-951-742-7
4	158.2	10.1	1777	1	US-08-173-508-5
5	158.2	10.1	1777	2	US-08-265-310-5
6	158.2	10.1	1777	3	US-08-951-742-5
7	113.2	7.3	1908	1	US-08-173-508-1
8	113.2	7.3	1908	2	US-08-265-310-1
9	113.2	7.3	1908	3	US-08-951-742-1
10	65.2	4.2	4257	2	US-08-690-473-1
11	65.2	4.2	12001	1	US-08-458-568A-11
12	52.6	3.4	2312	1	US-07-736-178C-1
13	49.2	3.2	44377	2	US-08-804-227C-7
14	49.2	3.2	44377	2	US-08-804-196-1
15	47.6	3.0	43280	2	US-07-804-227C-1
16	46.6	3.0	8438	1	US-07-945-283-1
17	45.2	2.9	30001	1	US-08-125-468-1
18	45.2	2.9	30001	2	US-08-474-933-1
19	45	2.9	12588	2	US-08-387-942C-1
20	45	2.9	20235	1	US-07-642-734C-3
21	45	2.9	20235	3	US-08-439-009A-3
22	44.4	2.8	1288	1	US-08-440-856A-9
23	44	2.8	38506	3	US-09-320-878-19
24	43.4	2.8	11219	1	US-07-642-734C-1
25	43.4	2.8	11219	3	US-08-439-009A-1
26	43	2.8	1167	1	US-07-960-985-1
27	43	2.8	1167	1	US-08-496-671-1
28	43	2.8	5970	3	US-09-320-878-21

C 29	42.4	2.7	1052	2	US-08-403-852D-10	Sequence 10, Appl
C 30	42.4	2.7	1052	3	US-08-510-646B-10	Sequence 10, Appl
31	42.4	2.7	2097	3	US-08-941-445A-10	Sequence 10, Appl
32	42.4	2.7	5467	2	US-07-745-206A-12	Sequence 12, Appl
33	42.4	2.7	5467	2	US-08-311-363-12	Sequence 12, Appl
34	42	2.7	1862	2	US-08-999-723-1	Sequence 1, Appl
35	42	2.7	1862	2	US-09-434-427-1	Sequence 1, Appl
C 36	42	2.7	43280	2	US-08-804-227C-1	Sequence 1, Appl
37	41.6	2.7	1659	3	US-09-083-351-3	Sequence 3, Appl
38	41.6	2.7	3946	3	US-09-083-351-1	Sequence 3, Appl
C 39	41.4	2.7	1833	2	US-08-403-852D-6	Sequence 1, Appl
C 40	41.4	2.7	1833	3	US-08-510-646B-6	Sequence 6, Appl
41	41.4	2.7	11219	1	US-07-642-734C-1	Sequence 1, Appl
42	41.4	2.7	11219	3	US-08-439-009A-1	Sequence 1, Appl
C 43	40.4	2.6	68750	3	US-09-335-409-1	Sequence 1, Appl
C 44	40.2	2.6	1269	3	US-08-396-218-1	Sequence 1, Appl
45	40.2	2.6	1269	1	US-08-760-116-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-08-173-508-7
Sequence 7, Application US/08173508
Patent No. 5616485
GENERAL INFORMATION:
APPLICANT: Bartfeld, Daniel
APPLICANT: Butler, Michael J.
APPLICANT: Hadary, Dany
APPLICANT: Jenish, David
APPLICANT: Krieger, Timothy
TITLE OF INVENTION: STREPTOMYCES PROTEASES AND IMPROVED
TITLE OF INVENTION: STREPTOMYCES STRAINS FOR EXPRESSION OF PEPTIDES AND
TITLE OF INVENTION: POLYPEPTIDES
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W.
CITY: Washington, D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/173,508
FILING DATE: 23-DEC-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 18740/125/CACO
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202 672 5300
TELEFAX: 202 672 5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1820 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 104..1720
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 104..244
FEATURE:
NAME/KEY: mat_peptide


```

1 CLASSIFICATION: 435
2
3 PRIOR APPLICATION DATA:
4 APPLICATION NUMBER: US 08/173,508
5 FILING DATE: 23-DEC-1993
6 ATTORNEY/AGENT INFORMATION:
7 NAME: BENT, Stephen A.
8 REGISTRATION NUMBER: 29,768
9 REFERENCE/DOCKET NUMBER: 18740/133/CACO
10 TELECOMMUNICATION INFORMATION:
11 TELEPHONE: 202 672 5300
12 TELEFAX: 202 672 5399
13
14 TEXEL: 904136
15
16 INFORMATION FOR SEQ ID NO: 7:
17
18 SEQUENCE CHARACTERISTICS:
19 LENGTH: 1821 base pairs
20 TYPE: nucleic acid
21 STRANDEDNESS: double
22 TOPOLOGY: linear
23
24 FEATURE:
25 NAME/KEY: CDS
26 LOCATION: 104..1720
27
28 FEATURE:
29 NAME/KEY: sig_peptide
30 LOCATION: 104..244
31
32 FEATURE:
33 NAME/KEY: mat_peptide
34 LOCATION: 245..1720
35
36 US-08-265-310-7

```

Query Match	17.0%	Score 265.2;	DB 2;	Length 1821;
Best Local Similarity	52.5%	Pred. NO. 1e-48;		
Matches 788; Conservative	0;	Mismatches 678;	Indels 36;	Gaps 8;

[illegible]

QY	671	agatcaactactactcgtgatacagctacagcccaaglttggcagccgttaacctcgaaacgct	710
Db	825	AGCTGACCTTACGTGGGGAGCGCTGTCGATGGGCACTTCTCGGGCCGGACCTACGGCGGTGTGT	884
QY	731	tcgtactcaatcgtcgtcgagatgctcctcgaagcgagctatcgatccagagcgtcttagcccaa	790
Db	885	TCGCCGACCGGACGGGGCCGCTGTCTCTGAGACGGCCGATGGACCCCTCGCTGTGCCGCC	944
QY	791	tcgaagaaagcatcagccaaatgcyggaatttcagaacgcttccaatgtactaagcccgccg	850
Db	945	GCCGCTCTGAACCTGGAGCAGACGGAGGGCTTCGAGACGGGGGTTCGAGTCCCTTCGCGAAGG	1004
QY	851	actcgcgccgctcgcgcgctgcctct-----ggcaccgactcgcgcgaagtgg	901
Db	1005	ACTCGGTGAAGCAACCCCGAGCTGGCCCCCTCGGCACAAAGGACACACCCCGACAGCTCG	1064
QY	902	tcaacgcctacacgaagccttggttgaaacccgtgtgtgaaagacccggttaagagctcgatc	961
Db	1065	GCAAGAACCTCAAGTCTTCTTGAGACACTTGAGCCGGAAGCCCTTGCCCCGCGCCACAG	1122
QY	962	caagctgc-----cttagctacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgtgtlaca	1015
Db	1125	CCGAGCGCCGCAAGCTCAGCCGAATCCCTCGCCACACACCGCGGATGCGCGGATGTACG	1184
QY	1016	gcctcgaagcgttgtaagtaacctcaagcgaatgtctgtcgtgagctgcgaagcgcgcgcg	1075
Db	1185	ACGAGGGGCGCTTGCGACACTCTGGCGAGTCCCTCACTCTGGCGATCAAGGAGGAAGAG	1244
QY	1076	ccgcgcgaactctcgtgtcgttcgcgcgaactctgaagcgcgcgcgcgcgcgcgcgcgcgc	1133
Db	1245	GTCGGGGCTGTCTATCTCTCCGACAGCTACTACAGACGCGGCGCGGCGCGCTGTACA	1304
QY	1136	gcaacacacaggaagcgcgttcaacgcgcgtccgcggtgcgtctatctgcgcacacacgcgc	1195
Db	1305	GCAACCTGATGTGTGCGCAAGCGCGCGCTGTAATGCTCGACCTCCCGCGCTTGTCT	1364
QY	1136	cagcgagcctcgtgtgc	1255
Db	1365	CCCCGGAGAGGTGTCCGGAGCGCCCTCCCGCACTTCAGAAAGGCGTCCCGGATCTTGGCG	1424
QY	1256	ggaagttcaacgcgcgtccgc	1315
Db	1425	AGGGCTTCGGTGGTCTCTCCCTGAC-----TGCCTGATCTGGCCGGTGAAGCCACAGG	1478
QY	1316	cgacgc	1375
Db	1479	GGGAGCGCACCGCATGAGAGCGGCGGCGCGCACCCCGATGCGTGGTGGCGCACACCC	1538
QY	1376	acgaacccgc	1433
Db	1539	GCGACCGCGGCCACCCCTTACCGGTGGGCGGAGGCGCTTCTCCAGCAAGCTCACCTCGGCC	1598
QY	1433	cgatgcataccttcgaagc	1492
Db	1599	ACCTTCCTACCTTACGAGGAGGAGCGGCACACCGCGTACGCGCGCGCACTCTCTCATCG	1658
QY	1493	actctgcgtgatcgaactattcttcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	1552
Db	1659	ACTCGCGGATCAACACTTACTCTGATCGCGGACCGCCCGGAGGAGCGCAAGCCCTGCT	1718
QY	1553	cg 1554	
Db	1719	CG 1720	

RESULT 3
US-08-951-742-7
; Sequence 7, Application US/08951742
; Patent No. 6127144
; GENERAL INFORMATION:
; APPLICANT: Barfield, Daniel
; APPLICANT: Michael J. Butler


```

Db      1539  GGGACCGGCGCACCCCTACCGCTGGGCGCGAGGCCCTCTCCAGCACAGCTCACCTCCGGCC 1598
OY      1433  cgcgtacaccccttcacgagccaccaacacacgcgtggttcgatgcaaccagtgltgg 1492
Db      1599  ACCCTCTCACCTACGACGAGGAGACGGCGACACCGCGTACGGCGGACGCTCTGCATCG 1658
OY      1493  actctcgagtgatgtacactatttcttcgaagggaacttcgcgcgaagatctcgagtcg 1552
Db      1659  ACTCGCGGATCAACAAGTACGTCGTACCGGACCGCGCCCGGAGGAGACGCAAGCGCTGCT 1718
OY      1553  cg 1554
Db      1719  CG 1720

RESULT 4
US-08-173-508-5
; Sequence 5, Application US/08173508
Patent No. 5616485
GENERAL INFORMATION:
APPLICANT: Bartfeld, Daniel
APPLICANT: Butler, Michael J.
APPLICANT: Hadary, Dany
APPLICANT: Jenish, David
APPLICANT: Krieger, Timothy
TITLE OF INVENTION: STREPTOMYCES PROTEASES AND IMPROVED
TITLE OF INVENTION: STREPTOMYCES STRAINS FOR EXPRESSION OF PEPTIDES
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W.
CITY: Washington, D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/173,508
FILING DATE: 23-DEC-1993
CLASSIFICATION: A35
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 18740/125/CACO
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202 672 5300
TELEFAX: 202 672 5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1777 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 190..1731
FEATURE:
NAME/KEY: misc_feature
LOCATION: 190..192
OTHER INFORMATION: /note="Met at position 1
OTHER INFORMATION: represents fMet"
US-08-173-508-5
Query Match 10.1%; Score 158.2; DB 1; Length 1777;
Best Local Similarity 49.7%; Pred. No. 8,4e-26;
Matches 663; Conservative 0; Mismatches 618; Indels 54; Gaps 8

```

QY	211	cgcttcgcgcgaagcgtgtctccgctcccccgtccagatccagccaaaccgggtgggagacaacagcgaag	270
Db	415	CGGTGCGCCACGACTGTAAGAGCAACCGCTGTGATGTGCTCGACCCGACGGGAGCAAGCATCGAT	474
QY	271	ctggcgcgtgatactccgctgcccccgcga---cggaagacagatctcgagacactgtgtgatacat	327
Db	475	CTGCGCGGTGATCGCGTCCCGGCGGAGGGGAGACGACCGCATCGGCTCCCTGCTGTTTAAC	534
QY	328	ccctggggagaccggggcgctctggcgcgttcagataatgtctgcgcgcctatgtgacccgcgatctgc	387
Db	535	TTTCGGCGGCGCCGGCGCGCTTCGGGCGGTTCACACATCC------GTCTACCCCGAACAC	588
QY	388	gaacacgcacatctccgcgaacttcagactcgtgtggcttgcttcagccagaagggctggccac	447
Db	589	GTCTCTCTCCCTGGCACAGACGGGTACAGACTGTGTAGCTGTGGGACCCCGCGGGGTGGCG--	646
QY	448	tcgaaccctcgtctgtcggtgtgcgaaccgagcccgagttcgaacgcgtataccgcgcgaatccg	507
Db	647	-CCAGCGAGGGCGTCCCGCTCCCGCACCGACGAGAGCGCATGAGAGCGCGCGAGTCCGTGTGAC	705
QY	508	atgctgcgaactcaatctccggcgccgcgtcaatcccaatctgaacacaggtctacccgcaagtctgc	567
Db	706	TTCCAGCGCGGACTTCGCCGGCGGAGAGACAGCGCTTACTTAAAGAGCGCCCGACTTTCGGC	765
QY	568	caggactctgttcgacgcgagatcggactcaagctctctcttgcgaataatcgtatacgcgtctc	627
Db	766	AGGGGCTCCGAGAA---GGCGCGCGGGAAGCATCATGTAAGAACGTGTGACGACGACGACAG	822
QY	628	gcacgcggagacatgagacatgtgtctgcgaagcgctatagtgtacgaatcagatacctaactctcga	687
Db	823	GCCCCGCAATGAGCACTGATGTCGGCACGTCTTGCGGAGCAGAGAGTGCATCTTACTTTCGGC	882
QY	688	tacagctacgcgcacaagatcttgggcaaccgcgtctaacctcggaaacggtctcgtatacatgttgcg	747
Db	883	ATTCCTCAACGGCACCGAATCTGGCGGGCGTGTACGCCCATCTGTTCGCCGAGCATGCTGGCG	942
QY	748	gcgaatgtctctcgaacgcgtcatalcgaaccgcgttatgcaccaatcgaagaaagacatcagc	807
Db	943	CGCGTGAATCTCGACGCGGCTGTGTGACACCGGCGCGCGACATGAGGACACCGCACCGAGAAC	10020
QY	808	caaatgycggagatlttcagaccgcttccaatagtataacgcgcgcagacttgcgcgcgtctgcgcg	867
Db	1003	CAGGCCAGGGGTTTTCAGACCGCGCGCTGTGAGCATCTTGAGTTCAGTCGCGCGCAGGAAACCC	10628
QY	868	gcccgcacctcgtgggcacagcagactcgcgcgcagttggttcaacgcgtataccagccctgtgtgac	927
Db	1063	GACACAGGGGTGCGGAAGATCGCGGCGCTG-----CTGGAGCGG	11010
QY	928	ccgcgtgtgtcagaagaccggglaaagacgttcgatacgaatcgaatcgttcgtacgtcgcgcgaacg	987
Db	1102	CTGGACGCGCGGACACTGCGCCACAGCTCCTCCCGCGGGGCGGAGACTGTAGCGACGACCTCCGCG	11610
QY	988	acgcgcgcgcacacatacaacgcgtgtatacagcccctcagcgtcttgaaagtacatctgacgcagtg	1047
Db	1162	TTTACCGGCACTGTGCTGCTGCGGCTTACAGAGAGCGGCTGGCGGCGCTCTACACGATGCG	12212
QY	1048	ctgcctcgtgtcgtcagc	1107
Db	1222	CTTGAAGGCGGCGCGAGGAGG---GCGACGCGTCTCGAATTTGCTGGCCCTCGCCAGCGGCTAC	12787
QY	1108	gacgcgcggggtgtgacg	1167
Db	1279	AACGAGCGCTGATCCCTCGGGGCGCTAGGACGACGACGACCCACTCGAAMAAGGTCTATATGCG	13388
QY	1168	ctgcgtctatgtgcccaacacacgc	1227
Db	1339	TGCGTGGACGACAGCGAGGCGCGACCGTGTGAGAGACAGAGAAGCTGCTGCGAGGTTTC	13988
QY	1228	cgtcaagctgcgcccgcttctctagtatacgggcagatltcaacgcgatctgcgcgcgcgcgcgcgcgc	1287
Db	1399	GAGAAGGCTCTCTCCGCTTTCGCGGCC-----TTCTCTCGGCTGGGACACGGCCGCGGTGG	14520

[illegible]

RESULT 6
 US-08-951-742-5
 : Sequence 5, Application US/08951742
 : Patent No. 6127144
 : GENERAL INFORMATION:
 : APPLICANT: Bartfeld, Daniel
 : APPLICANT: Michael J. Butler
 : APPLICANT: Dany Hadary
 : APPLICANT: David Jenish
 : APPLICANT: Tim Krieger
 : APPLICANT: Lawrence T. Malek
 : APPLICANT: Gisela Soostmeyer
 : APPLICANT: Eva Walczyk
 : APPLICANT: Phyllis Kirgsmann
 : APPLICANT: Shella Garven
 : TITLE OF INVENTION: METHOD FOR EXPRESSION OF PROTEINS IN
 : TITLE OF INVENTION: BACTERIAL HOST CELLS
 : NUMBER OF SEQUENCES: 25
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Foley & Lardner
 : STREET: 3000 K Street, N.W., Suite 500
 : CITY: Washington
 : STATE: D.C.
 : COUNTRY: USA
 : ZIP: 20007-5109
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: Patentin Release #1.0, Version #1.30
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/951,742
 : FILING DATE: 16-OCT-1997
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Bent, Stephen A.
 : REGISTRATION NUMBER: 29,768
 : REFERENCE/DOCKET NUMBER: 0189740/0140
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: (202)672-5300

```

? TELEFAX: (202)672-5399
? TELEX: 904136
? INFORMATION FOR SEQ ID NO: 5:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1777 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: double
? TOPOLOGY: linear
? MOLECULE TYPE: cDNA
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 190..1728
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: 190..192
? OTHER INFORMATION: /product="Met at position 1
? OTHER INFORMATION: represents fmet"
US-08-951-742-5

```

	Query Match	10.1%	Score 158.2	DB 3:	Length 1777;	
	Best Local Similarity	49.7%	Pred. No. 8.4e-26;			
	Matches 663; Conservative	0;	Mismatches 618;	Indels 54;	Gaps	8
QY	cgtctgcgacgaagtgatccgcccgcgtcgacatacggacaaccacccggttggacacaaagcgaag	270				
Dd						
QY	415 CGGTGCGCACCGCTTAAGGCACCGCTGGACTGTCTCCGACCCTCCGACGGCAGACGATCGMT	474				
QY	271 ttgacggtgatccgcgcgtccgcga---cggacagcgaatttcggacactcgttgtcaaat	327				
Dd						
QY	475 CTCGGCGTGATCCGGTTCGCCGGCGGAGCGGGGAGACCGCATGCGGCTCCCTCGTGTTCAAC	534				
Dd						
QY	328 ccttgaggagacccgggggcgcgcgcgtcgacatacgtgcgcgcataatggaccacgcgatccgc	387				
Dd						
QY	535 TTTCGGCGCCCCGGGGCGCTCTCCGGCGTTCACAAGATGCC-----GTCCTAACCGCCACACC	588				
Dd						
QY	388 gacacacgaacattctccgcacacttcgacactcgtgtagcgtcttcgacccgaagaagtctgacac	447				
Dd						
QY	589 GTCCTCTCCTCTCACAGAGAGGGATTACGACCTGGTAGAGCTGGGACCCCGGGGGGTGGCG--	646				
Dd						
QY	418 tcgacccctcgtctgcgtgtcgcgaacgcgcgaagtctgacgcgttaacgcgcgcgaatccg	507				
Dd						
QY	647 -CCACCGAGGGGCGTCCGGCTCGCCGACCGCAGAGCGCATGAGAAGCCCGCCAGTCGATGGAGAC	705				
Dd						
QY	508 atgycgcgaataagtcgcgcgcgcgcgcgtccaaccaacgttcgaacaggtctaacgcgcagttgcc	567				
Dd						
QY	706 TTCACGCGCGAGATCCCGCGCGCGAGGACGAGCGCTTACTGTAAGAGACCCCCTCCGACTTCCGC	765				
Dd						
QY	568 cagaagcttgtctgcacgcgaatgaggtcttaacgtctcttcgtgccaatatcgtlaacgcgcttcgc	627				
Dd						
QY	823 GCCCCGCAATGACCTGATGCGGACGCTCCGCGGCGAGAGAGATGACATACTTCGCCG	882				
Dd						
QY	688 tacagctacgcgaaccaaagttagaacacgcgttacccugaaacggttcgttaactaagtgcgag	747				
Dd						
QY	883 ATCTCTTACGACACCGAAGACTCGCGCGCTCTACGCCACTCTGTTCCCGCAGCACGTGGGCG	942				
Dd						
QY	748 gcgaatgtctctcgaacgcgtctaactcacgcgcgttaagcccaatcgaagaaaacalcaga	807				
Dd						
QY	943 CCGCGATCTCTGACGCGCGGTGTGGACCCGGGGCGCGACACGATGGGCCACGCGCGAGAAC	1002				
Dd						
QY	808 caaatgycgaatcttcagaacgcgtcttaactgaactacgcgcgcgcgcgcgcgcgcgcgcgcgc	867				
Dd						
QY	1003 CAGGCCAAGGGGTTTTCCAGGCGCGCGCTGGACGACTACTGAGTCAGCCGCGCAGGAACCC	1062				
Dd						
QY	868 gccctgcctctcgtgcacacgc	927				
Dd						
QY	1063 GAAACAGGGGTCCGGAAAGATCCCGCGCTG-----CTGGACCGG	1101				
Dd						
QY	928 ccgcctcgtgtcagaagaacgcggtlaagacgttcgcatccacgtgcgtgacctgacgcgcgcgcgcgc	987				
Dd						

Db 1102 CTGGACCCGAGCCACTGCCACCTCTCCGCCGCGGAGGTGACGACAGCCCTGCGC 1161
Qy 988 acgagcgagccaccatcaacgcgctgtacagccctacgcgctggaagtacccctgacagcgtgt 1047
Db 1162 TTCCACCGGACATCGCTGCGCCGTGTACAGAGAGCGCGCTGGCGGCTGACCAAGTGGC 1221
Qy 1048 ctgctgagcgctgcagcgcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagc 1107
Db 1222 CTGAAGGCGGCGGAGAGG---GCGAGCGGCTCGAGATTGGTGGCCCTCGCGGAGCGGCTAC 1278
Qy 1108 gaegcgcgagatgcagagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgt 1167
Db 1279 AACGAGCGGTATCCCTCGCGGCGGTACGAGCAGACCACTCGCAAAAGGCTATATCG 1338
Qy 1168 tgcgtcatgcgcccacacgcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagc 1227
Db 1339 TGCCGTGACACAGACAGAGAGCGCCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGTTC 1398
Qy 1228 cgtcaggtcgccccgtctcctagctacgagcgagcgagcgagcgagcgagcgagcgagcgt 1287
Db 1399 GAGAGAGTCTCTCCCTCTCTCGCGCC-----TTCTCGGCTGGAGACGCGCGGCGGTTGG 1452
Qy 1288 tgcgagctgtgcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagc 1347
Db 1453 TGCCACGACATGCGCGGTGGCGGTACAGACAGCGAGAGAGAGAGAGAGAGAGAGAGAGAG 1512
Qy 1348 ggcgaagtcgt 1407
Db 1513 GCGCGGCTCTGCTGT 1572
Qy 1408 gacctgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagc 1458
Db 1573 AGGATGGCGGAGCGAGCTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1632
Qy 1459 caacatgcgagtggt 1518
Db 1633 CACGCTGCTACAGGAG 1692
Qy 1519 gacgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagc 1533
Db 1693 AAGGAGCGGTGCCG 1707

RESULT 7
US-08-173-508-1
Sequence 1, Application US/08173508
Patent No. 5616485
GENERAL INFORMATION:
APPLICANT: Bartfield, Daniel
APPLICANT: Butler, Michael J.
APPLICANT: Hadary, Dany
APPLICANT: Krieger, Timothy
TITLE OF INVENTION: STREPTOMYCES PROTEASES AND IMPROVED
TITLE OF INVENTION: STREPTOMYCES STRAINS FOR EXPRESSION OF PEPTIDES AND
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W.
CITY: Washington, D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/173,508
FILING DATE: 23-DEC-1993
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 18740/125/CACO
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202 672 5300
TELEFAX: 202 672 5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1908 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 146..1759
FEATURE:
NAME/KEY: misc-feature
LOCATION: 146..148
OTHER INFORMATION: /note="Met at position -39"
OTHER INFORMATION: represents fmet"
FEATURE:
NAME/KEY: sig-peptide
LOCATION: 146..262
FEATURE:
NAME/KEY: mat-peptide
LOCATION: 263..1756
US-08-173-508-1

Query Match 7.3%; Score 113.2; DB 1; Length 1908;
Best Local Similarity 46.1%; Pred. No. 3.7e-16;
Matches 669; Conservative 0; Mismatches 743; Indels 40; Gaps 7;

Qy 12 ggcgcgagacgagtgagctgcgagcgagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 71
Db 197 GCATGTGTACCCGACGACGATGCGCGCGCGCTTCGCGACCGCGCGGAGAGCGCGCC 256
Qy 72 cctagcagagccccgttgcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagc 131
Db 257 CCGGCGCAGCGGCGGCGGCGGCGGAGCTGGAGCGGAGCGCGCGCGCGCGCGCGCGCGCGCG 316
Qy 132 tgcgcgagctgtgcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagc 191
Db 317 GCG 372
Qy 192 cagcgaaatcgcagctgcagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagc 251
Db 373 GAACCTGCCCAAGCCCATCCAGTGGCGGCTACGTCACGCGTCCGATGACGACCAAGCC 432
Qy 252 cgtgtgagacaaagcgagtggt 302
Db 433 GTACGGCAAGCAATCATGCTGCG 492
Qy 303 gcatcgcagagcgagtggt 362
Db 493 GCGCGAGGCG 552
Qy 363 cgcgcgtatgcagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagc 419
Db 553 GCG 612
Qy 420 ggcgttcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagc 479
Db 613 CGGCTTCGACCG 672
Qy 480 cgcgttcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagc 539
Db 673 GTTCGTAAAGCAACGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAG 732
Qy 540 cgtcgaacgagctacccgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagc 599

OY	192	cagggaataatcgcacgtcgaacgctcgcgcgaacggtgcgtcccgctccgcttcgaatacgaacac	251
Db	373	GAACCTGCCCAAGCCCATTCAGTGGCGGCTACAGGTGCGGATGAGTACGACCAAGCC	432
OY	252	cagttgtagcaacaagcgaagtttggcgtgtatccgcgtcccgcgagcgga-----ca	302
Db	433	GTACGGCAAGAGATCTAGGCTTCGCCGTCGACCCGATCGGCAACACCGGAACCGAGAGCGA	492
OY	303	gcgaattcggagcaatcgtctgtcaatccctcggggagaccggggcgctcggcgttcgaatagt	362
Db	493	GCGCGAAGGGGCCCCGTGATCTACAAACCCCGCGGTCCCGGCGGCTCCGGCTCCGTTTCC	552
OY	363	cgcgcgtatggaacccgcgcgtccgcac-----acgaatacttcgcgaacttcgaactgt	419
Db	553	GGCCCCGCTCAGCAAGAAAGCCCGGCTGTGGGCAACAGGCGCAAGGCTTACGACTTGT	612
OY	420	ggcgttcgaaccgagagggttcggccaactcgaaccctcgttgcgtgtcgcacgcagc	479
Db	613	CGGCTGTGACCCCGCGCGGCTCGGCGCACTCCGCGCCCACTCTCGTGTCGACCCGACGA	672
OY	480	cgaattcgaacggttacccggtcgatccgaatcgaatggcgacatacagttccgcgcggtccaca	539
Db	673	GTTCGTCAAGGCACCCAAAGCCCAACCCCGGCCGCTCCGAGCGCAAGCGGCGCA	732
OY	540	cgttcgaacaggtctcaacgcgaagtttggccgaagacttgttgaacgcgaatggttcacgt	599
Db	733	GCGC---AAGTCGCCCCGCGAGTACGCCGAGGCGGCTCTTCGAGCGGAGGCGAGATGCT	789
OY	600	cttggccaatatcgtgtacgcggtccgtccgttcgaacgggacatgtgacatgtgttcgcaagcgt	659
Db	790	CCCCACATGAC---ACGCGCAACACCGCGCGCGACCTCGACGTATCCGGCGGCGCT	846
OY	660	aggtgcagatcaagaatcaactactcgtatacagatacgcgcacaaagtttggaccgccta	719
Db	847	CGGGGAGAAAGAGTCAACTACCTCGGCGTCTCTACGGGACCTACCTACGCGCGCTA	906
OY	720	ccgtggaacggttcgtaactatgttcgggcgcatgttccttcgaacggcgctatcatcaacg	779
Db	907	CGGCACCTCTTCCCGGACACGCTCGCGGCAAGTGTCTGACAGCGTGTCAACCCGTC	966
OY	780	cgtttagcccaatcgaaggaagaagcaatcgaacaaatggcggaatttcgaacgcgttcaatga	839
Db	967	CCGGACACAAGATCTGTATCCAGGCGCAAC-----CTGGACACAGAGCTGTGC	1011
OY	840	ctacgcgcgcgaactgcgcgcgcgtccgcgcgcgtccctcctcgtgcacacgaactcgcgcagtg	899
Db	1012	CTTGAGAGGCGCGGTGGAAGGACTGTGGAGCTGGCGCGGAGACGACGCGGCTTACCA	1071
OY	900	ggttaaacgcgtctacacgcgccttggttgaaccgcgtgtgtgcagaagaagccgggttaagaagtcg	959
Db	1072	CTTCGGGAGACACCCGCCCCAGGCTCAAGACCAAGTGTGTGAAGCTGCGCGCCGCGCGC	1131
OY	960	tccacgtgcgccttagctacgcgcgcgcgaacgaaggaacacatcaacgcgcgttgtacaccc	1019
Db	1132	GAAGAAACCGCTGGGCGGCGCTGTCGAGACCGCGCGAGCTGTGATCTCTTCCAGAGGCG	1191
OY	1020	tcaagcgttcggaagtacctgcgcgcagtggtctcgtcgtgggtcgcgcgcgcgcgcgcgcgcg	1079
Db	1192	CCCCCTACTACGACTCCCTCTGGGCGCGACCGCGGAGATCTTACACAAAGTACGTGCCGG	1251
OY	1080	cgaacttgctgtcttbcgcgaacatatga---cgacgcggatgtaagaacgggcgaactacag	1136
Db	1252	CGAACCCAGCGGCTGCTCGACCGCGCGCACCCGACCTGTCCGACACCGCGGGAACGC	1311
OY	1137	caacgcacgaagcgcgttcaacgcggttcgcgtgtcgtatcgtcgcgcacaacgcgcgcgtacc	1196
Db	1312	CTTCGCGGAAGACGGCAACGCCCTCTACACAGGCGCTGTGAGTGCACACGAGCCAAATGTGC	1371
OY	1197	agcgcttcgttggtgcgcgcgaacaaagatcgcgtgcgtgcgcgcgcgcgttccttagtagtg	1266
Db	1372	CGCCAACTGGGCACTCGGGAACCGGGAACAACCCGGGCTCAACCGCGACCAACCGGTTAT	1431

[illegible]

```

1 RESULT 9
2 US-08-951-742-1
3 : Sequence 1, Application US/08951742
4 : Patent No. 6127144
5 : GENERAL INFORMATION:
6 : APPLICANT: Bartfeld, Daniel
7 : APPLICANT: Michael J. Butler
8 : APPLICANT: Dany Hadary
9 : APPLICANT: David Jenish
10 : APPLICANT: Tim Krieger
11 : APPLICANT: Lawrence T. Malek
12 : APPLICANT: Gisela Soostmeyer
13 : APPLICANT: Eva Waliczky
14 : APPLICANT: Phyllis Krzyzsan
15 : APPLICANT: Sheila Garven
16 : TITLE OF INVENTION: METHOD FOR EXPRESSION OF PROTEINS IN
17 : TITLE OF INVENTION: BACTERIAL HOST CELLS
18 : NUMBER OF SEQUENCES: 25
19 : CORRESPONDENCE ADDRESSES:
20 : ADDRESSEE: Foley & Lardner
21 : STREET: 3000 K Street, N.W., Suite 500
22 : CITY: Washington
23 : STATE: D.C.
24 : COUNTRY: USA
25 : ZIP: 20007-5109
26 : COMPUTER READABLE FORM:
27 : MEDIUM TYPE: Floppy disk
28 : COMPUTER: IBM PC compatible
29 : OPERATING SYSTEM: PC-DOS/MS-DOS
30 : SOFTWARE: Patentin Release #1.0, Version #1.30
31 : CURRENT APPLICATION DATA:
32 : APPLICATION NUMBER: US/08/951.742
33 : FILING DATE: 16-OCT-1997
34 : ATTORNEY/AGENT INFORMATION:
35 : NAME: Bent, Stephen A.
36 : REGISTRATION NUMBER: 29,768
37 : REFERENCE/DOCKET NUMBER: 0189740/0140
38 : TELECOMMUNICATION INFORMATION:
39 : TELEPHONE: (202)672-5300
40 : TELEFAX: (202)672-5399
41 : TELEX: 904136
42 : INFORMATION FOR SEQ ID NO: 1:
43 : SEQUENCE CHARACTERISTICS:
44 : LENGTH: 1908 base pairs
45 : TYPE: nucleic acid
46 : STRANDEDNESS: double
47 : TOPOLOGY: linear
48 : MOLECULE TYPE: cDNA
49 : FEATURE:
50 : NAME/KEY: CDS
51 : LOCATION: 146..1756
52 : FEATURE:
53 : NAME/KEY: misc_feature
54 : LOCATION: 146..148
55 : OTHER INFORMATION: /product= "Met at position -39
56 : OTHER INFORMATION: /represents fmet"

```


C	13	50	3.2	24379	18	T93095	Streptomyces freno
C	14	50	3.2	24379	19	V59925	Streptomyces roseo
C	15	49.2	3.2	44377	18	T78508	Platenoides synthia
C	16	49.2	3.2	44377	18	T80414	Platenoides synthia
C	17	48.4	3.1	26338	19	V62134	HSV-2 strain SB5 C
C	18	47.6	3.1	15872	18	T68715	Streptomyces venez
C	19	47.6	3.1	15872	21	T87283	S. venezuelae vep
C	20	47.6	3.1	43280	18	T80413	Tylosine synthase
C	21	46.8	3.0	3222	19	V62382	B. subtilis ECB de
C	22	46.8	3.0	4050	19	V62379	B. subtilis ECB de
C	23	46.8	3.0	4093	19	V62381	B. subtilis ECB de
C	24	46.6	3.0	8438	15	O73500	DNA encoding pseud
C	25	45.2	2.9	30001	18	T61016	Total DNA sequenc
C	26	45.2	2.9	30001	20	X05110	S. aureofaciens DN
C	27	45	2.9	12588	15	O63293	Sequence encoding
C	28	44.8	2.9	77536	21	A14651	Nucleotide sequenc
C	29	44.2	2.8	53789	19	V21187	Amlycolatopsis med
C	30	44	2.8	10732	21	A10594	Gene encoding a su
C	31	44	2.8	11220	21	T87298	S. venezuelae macr
C	32	44	2.8	36778	21	T87318	S. venezuelae mact
C	33	44	2.8	37948	21	T87285	S. venezuelae pik
C	34	44	2.8	38506	21	Z56001	Recombinant cosmid
C	35	43.8	2.8	13842	21	T87297	S. venezuelae mact
C	36	43.6	2.8	2929	19	V62384	B. subtilis wild-t
C	37	43.6	2.8	3336	13	O23759	Gene encoding acut
C	38	43.4	2.8	29879	14	O46806	etuya region of S.
C	39	43.2	2.8	1976	20	T28095	Human PREA3 prote
C	40	43.2	2.8	2614	11	O06631	Gene conferring te
C	41	43.2	2.8	77536	21	A14651	Nucleotide sequenc
C	42	43	2.8	1167	14	O41760	Oxytocin receptor
C	43	43	2.8	1167	19	T76908	Human oxytocin rec
C	44	43	2.8	1331	21	A37128	Human POU1f81 (UNO
C	45	43	2.8	2787	21	T87302	S. venezuelae macr

ALIGNMENTS

RESULT	1	
ID	Q99366	standard; DNA; 1821 BP.
XX	Q99366;	
AC	Q99366;	
XX	04-DEC-1995	(first entry)
DT		
XX	S. lividans	protease P5-6 gene.
DE		
XX	Protease; metalloendoproteinas	; tripeptidyl aminopeptidase;
KW	protease-deficiency; protein secretion;	ds.
XX		
OS	Streptomyces lividans.	
XX		
FH	Key	Location/Qualifiers
FT	CDS	104..1723
FT		/tag= a
FT	sig_peptide	104..244
FT		/tag= b
FT	mat_peptide	245..1720
FT		/tag= c
FT		/product= P5-6.
XX		
PN	WO9517512-A.	
XX		
PD	29-JUN-1995.	
XX		
PF	22-DEC-1994;	94MO-US14772.
XX		
PR	23-DEC-1993;	93US-0173508.
XX		
PA	(CANG-) CANGENE CORP.	
XX		
I1	Barfield D, Butler MJ, Hadary D,	Jenish DL, Krieger TJ;

[illegible]

Db	1573	aggatgagcgaagcgcgtggaacgaagcgttcgctgtgtcgtaccttgcagggcgaaggga	16322
Qy	1459	cacactcgcggtgttcgaacacacagttgtgtgactctgcggtgatgaactattctc	1518
Db	1633	cacggtgtcctacgggaacggaagcagctgtgtcgtacctcgcggtggaagcctactgttg	16922
Qy	1519	gacgggacacttgcgcg	1533
Db	1693	aaaggagacgltgcgcg	1707
RESULT	4		
Q99368			
ID	Q99368	standard; DNA; 1777 BP.	
XX	Q99368;		
AC	04-DEC-1995	(first entry)	
DT			
XX	S. lividans protease P8-2 gene.		
XX			
XX	Protease: metalloendoproteinase; tripeptidyl aminopeptidase;		
XX	protease-deficiency; protein secretion; ds.		
OS	Streptomyces lividans.		
XX			
FH	Key	Location/Qualifiers	
FT	CDS	88..1502	
FT		/*tag= a	
XX	W09517512-A.		
XX			
PD	29-JUN-1995.		
XX			
EF	22-DEC-1994;	94MO-US14772.	
XX			
PR	23-DEC-1993;	93US-0173508.	
XX			
PA	(CANG-) CANGENE CORP.		
XX			
PI	Bartfeld D, Butler MJ, Hadary D, Jenish DL, Krieger TJ;		
PI	Malek LT, Soostmeyer G, Walczyk E;		
XX			
DR	WPI: 1995-240673/31.		
DR	P-PSDB: R80508.		
XX			
XX	Endogenous Streptomyces protease(s), opt. having impaired activity -		
XX	useful in prodn. of exogenous proteins with reduced proteolytic		
XX	degradation		
XX	Claim 7; Fig.19: 142pp; English.		
XX			
CC	Protease Tap-negative cells were transformed with a S. lividans 66		
CC	genomic library and screened with APA-beta-naphthylamide to		
CC	isolate colonies contg. genes (Q99365-68) for novel proteases		
CC	P5-4, P5-6, P5-10 and P8-2 (R80505-08). Impaired expression of		
CC	such proteases by Streptomyces hosts improves the quality, quantity		
CC	and stability of exogenous gene products.		
XX			
XX	Sequence 1777 BP: 288 A; 607 C; 666 G; 216 T; 0 other;		

[illegible][illegible]

[illegible]

PE 16-APR-1998; 98WO-US07573.
 XX
 PR 16-APR-1997; 97US-0843659.
 XX
 PA (ARCH-) ARCH DEV CORP.
 XX
 PI Leopardi R, Roizman B;
 XX
 DR WPI; 1998-594559/50.
 DR P-PSDB; W80810.
 XX
 PT Use of herpes simplex virus *U(s)3* polypeptide - for developing
 PT products for modulating apoptosis in cells and for identifying
 PT compounds which act as stimulators or inhibitors of apoptosis
 XX
 PS Example 2; Pages 60-63; 85pp; English.
 XX
 CC This is the nucleotide sequence of Herpes simplex virus ICP4 used in
 CC the method of the invention as modulators of apoptosis. The methods
 CC and products can be used to identify compounds which modulate
 CC (stimulate or inhibit) apoptosis in cells. They can be used to
 CC immortalise cells for the study of these cells or for growing cells
 CC in large numbers for the productions of proteins. They can also be
 CC used for stimulating apoptosis in cells, e.g. for treating a subject
 CC with a HSV infection.
 XX
 SO Sequence 4257 BP; 412 A; 1768 C; 1663 G; 414 T; 0 other;

[illegible]

```

RESULT      8
V10362
ID   V10362 standard; cDNA; 4257 BP.
XX
XX   V10362;
XX
XX   30-JUN-1998 (first entry)
XX
XX   Infected cell protein number 4 alpha-4 gene.
XX
XX   Infected cell protein number 4; ICP4; alpha-4; cell apoptosis;
XX   therapeutics; ss.
XX
XX   Herpes simplex virus.
XX
XX   Key Location/Qualifiers
XX   CDS 361..4257
XX       /*tag= a
XX       /product= "Infected cell protein"
XX
XX   WO9804709-A2.
XX
XX   05-FEB-1998.
XX
XX   23-JUL-1997; 97MO-US12904.
XX
XX   26-JUL-1996; 96US-0690473.
XX
XX   (ARCH-) ARCH DEV CORP.
XX
XX   Leopardi R, Roizman B;
XX
XX   WPI, 1998-130697/12.
XX
XX   P-PSDB; W40200.
XX
XX   Use of herpes simplex virus ICP4 polypeptide - useful for, e.g.
XX   blocking apoptosis in cells, production of proteins and gene therapy
XX
XX   Disclosure; Fig 2; 63pp; English.
XX
XX   The herpes simplex virus (HSV) alpha-4 gene encodes infected cell protein
XX   number 4 (ICP4). Both the alpha-4 gene and ICP4 can be used to block
XX   cell apoptosis. Similarly the administration of an agent that inhibits
XX   ICP4 or the alpha-4 gene can induce apoptosis in HSV infected cells.
XX   This can be used for the immortalisation of cells, production of
XX   proteins, gene therapy, or inhibition of cell death induced in vivo.
XX   They can also be used for production of therapeutics comprising
XX   inhibitors of HSV ICP4 function, useful for treating HSV function.
XX
XX   Sequence 4257 BP; 412 A; 1768 C; 1663 G; 414 T; 0 other:
SQ
Query Match      4.2%; Score 65.2; DB 19; Length 4257;
Best Local Similarity 44.9%; Pred. No. 8.9e-05;
Matches 247; Conservative 0; Mismatches 303; Indels 0; Gaps 0;

```

```

QY 1083 ctgtctgtgtctgtccagactatgacgcccggatgacagcggcactacagcaaga 1142
DB 2595 cccgaagacgaagaagaagacgagcgccgagcccccgtcgtgacgcccgcctccc 2654
QY 1143 ccaagacgcttcaagcgggtccgtgtcgtctatgcccacacccgacgacagcgc 1202
DB 2655 cgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 2714
QY 1203 ctgggtggccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1262
DB 2715 gcccgcgccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 2774
QY 1263 caccggtatccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1322
DB 2775 cgaagggccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 2834
QY 1323 gcacccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1382
DB 2835 ggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 2894
QY 1383 ggcacactccg 1392
DB 2895 ggaccaccgcg 2904

RESULT      9
Q76213/c
ID   Q76213 standard; DNA; 12001 BP.
XX
XX   Q76213;
XX
XX   04-AUG-1995 (first entry)
XX
XX   HSV L/ST region.
XX
XX   HSV; junction-spanning transcript; L/ST; therapy; virucide; ds.
XX
XX   Herpes simplex virus type 1.
XX
XX   Key Location/Qualifiers
XX   TATA_signal 199..204
XX       /*tag= a
XX   misc_binding 220..225
XX       /*tag= b
XX       /note= "ICP4 binding site"
XX   misc_RNA 229
XX       /*tag= c
XX       /note= "5' end of the L/STs"
XX   misc_RNA 370..372
XX       /*tag= d
XX       /note= "first codon of ORF-1"
XX
XX   WO9428156-A.
XX
XX   08-DEC-1994.
XX
XX   20-MAY-1994; 94MO-US05770.
XX
XX   20-MAY-1993; 93US-0065146.
XX
XX   (DAND ) DANA FARBER CANCER INST INC.
XX
XX   Schaffer PA, Yeh L;
XX
XX   WPI; 1995-022825/03.
XX
XX   Herpes Simplex Virus (HSV) specific junction spanning transcript
XX   - for inhibiting HSV L/ST synthesis, in the treatment of HSV
XX   infection.
XX
XX   Disclosure; Page 38-44; 64pp; English.
XX
XX   An HSV-specific junction-spanning transcript (L/ST) maps at the 5'

```


[illegible]

Search completed: January 12, 2001, 20:05:55
Job time: 26239 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 12, 2001, 19:59:14 ; Search time 6876.06 Seconds
(without alignments)
1161.085 Million cell updates/sec

Title: US-09-461-774-3
Perfect score: 1560
Sequence: 1 gcggccatggtgcgcgcag.....gtctgcggtgcgcgcctga 1560

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 118133 seqs, 2558875100 residues
Total number of hits satisfying chosen parameters: 2236266

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_om:*
4: gb_ov:*
5: gb_ph:*
6: gb_pl1:*
7: gb_pl2:*
8: gb_pr1:*
9: gb_pr2:*
10: gb_pr3:*
11: gb_ro:*
12: gb_sy:*
13: gb_un:*
14: em_fun:*
15: em_hum1:*
16: em_hum2:*
17: em_in:*
18: em_om:*
19: em_or:*
20: em_ov:*
21: em_pat:*
22: em_ph:*
23: em_pl:*
24: em_ro:*
25: em_sts:*
26: em_sy:*
27: em_un:*
28: em_v1:*
29: gb_hc91:*
30: gb_hc92:*
31: gb_in1:*
32: gb_in2:*
33: em_ba1:*
34: em_ba2:*
35: em_hum3:*
36: em_hum4:*
37: gb_pr4:*
38: gb_hc93:*
39: gb_hc94:*
40: gb_hc95:*
41: gb_hc96:*
42: gb_hc97:*
43: em_hc91:*

44: em_hc92:*
45: em_hc93:*
46: em_hum5:*
47: gb_pl3:*
48: gb_pr5:*
49: gb_hc98:*
50: gb_hc99:*
51: gb_hc910:*
52: gb_hc911:*
53: gb_hc912:*
54: gb_hc913:*
55: gb_hc914:*
56: gb_in3:*
57: gb_hc915:*
58: gb_hc916:*
59: gb_hc917:*
60: em_hc94:*
61: em_hc95:*
62: em_hc96:*
63: em_hc97:*
64: em_hum6:*
65: gb_hc918:*
66: gb_hc919:*
67: gb_hc920:*
68: gb_hc921:*
69: gb_hc922:*
70: gb_hc923:*
71: gb_v11:*
72: gb_v12:*
73: gb_ba3:*
74: em_hc98:*
75: em_hc99:*
76: em_hc910:*
77: gb_pr6:*
78: gb_pr7:*
79: gb_sts1:*
80: gb_sts2:*
81: gb_pat1:*
82: gb_pat2:*
83: em_hc90:*
84: gb_hc924:*
85: gb_pr8:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	1553.6	99.6	38110	2	MTCY427
2	908.6	58.2	36404	2	MSGB32CS
3	265.2	17.0	1820	73	STMSLPD
4	265.2	17.0	1820	81	I38962
5	265.2	17.0	1821	81	AR027260
C 6	265.2	17.0	40544	73	SC5H
C 7	187.4	12.0	38366	73	SC23B6
C 8	185.2	11.9	41800	73	SCDBA
9	158.2	10.1	1777	81	AR027259
10	158.2	10.1	1777	81	I38961
11	158.2	10.1	1973	73	STMSLPE
C 12	134.6	8.6	36583	73	SC5H1
C 13	127.8	7.3	31360	73	SC9C7
C 14	113.2	7.3	1908	81	AR027257
15	113.2	7.3	1908	81	I38959
16	113.2	7.3	2295	73	STMTRLP
C 17	105.4	6.8	26555	73	SC2G1
C 18	79.4	5.1	14866	73	SC1B5
C 19	69	4.4	123580	1	AF263912
C 20	66.2	4.2	4583	73	SMASRUABC
21	66	4.2	38634	73	SC26C5

22	65.2	4.2	6633	71	HEHSV163	X06461	Herpes simp
C 23	65.2	4.2	12001	81	AR048721	AR048721	Sequence
24	65.2	4.2	26245	82	HS1US	L00036	HSV1 (Stral
C 25	65.2	4.2	26245	72	HS1US	L00036	HSV1 (Stral
26	65.2	4.2	152261	71	HE1G	X14112	Herpes simp
C 27	65.2	4.2	152261	71	HE1G	X14112	Herpes simp
28	62.6	4.0	8113	72	HSB1CIP4A	L14330	Bovine herp
C 29	62.6	4.0	135301	71	BHV1G1GEN	AJ004801	Bovine he
C 30	62.6	4.0	135301	71	BHV1G1GEN	AJ004801	Bovine he
31	61	3.9	34766	73	SCC121	AL137166	Streptomy
C 32	59.2	3.8	10798	1	AE001949	AE001949	Deinococc
33	58.6	3.8	41622	73	SCD25	AL118514	Streptomy
C 34	57	3.7	20394	73	SNM132222	AJ132222	Streptomy
35	56.2	3.6	2764	72	HSB1CIP4B	L14321	Bovine herp
C 36	55.8	3.6	5635	1	AF145039	AF145039	Streptomy
C 37	55.6	3.6	16185	73	SC985	AL139164	Streptomy
38	54.8	3.5	36368	73	SC985	AL139164	Streptomy
C 39	54.6	3.5	1235	1	AF187997	AF187997	Halobacte
C 40	54.6	3.5	2764	72	HSB1CIP4B	L14321	Bovine herp
41	54.6	3.5	12928	1	AE005076	AE005076	Halobacte
42	54.2	3.5	2028	47	SSI132828	AJ132828	Spermatoz
43	54.2	3.5	22789	73	SC2A11	AL031184	Streptomy
44	54	3.5	3986	73	SPWMSDA	L48580	Streptomy
45	54	3.5	69350	2	MTV004	AL009198	Mycobacte

ALIGNMENTS

RESULT 1
MTCY427/c 38110 bp DNA BCT 24-JUN-1999
Mycobacterium tuberculosis H37Rv complete genome; segment 99/162.
270692.1 AL123456
270692.1 GI:3261567

LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Mycobacterium tuberculosis.
Mycobacterium tuberculosis
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacteriaceae; Mycobacteriaceae;

REFERENCE
AUTHORS
1 (bases 1 to 38110)

Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C.,
Harris, D., Gordon, S.V., Eigmeier, K., Gas, S., Barry III, C.E.,
Tekala, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T.,
Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S.,
Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J.,
Moule, S., Murphy, L., Oliver, S., Osborne, J., Quail, M.A.,
Rajandream, M.A., Rogers, J., Rutter, S., Seeger, K., Skelton, S.,
Squares, S., Squares, R., Sultson, J.E., Taylor, K., Whitehead, S., and
Barrell, B.G.

TITLE
JOURNAL
MEDLINE
REMARK
2 (bases 1 to 38110)

Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence
Nature 393 (6685), 537-544 (1998)

ERRATUM: [[published erratum appears in Nature 1998 Nov
12:396(6707):1901]]
2 (bases 1 to 38110)

COMMENT
Direct Submission
Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium
tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome
Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique
Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux,
75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk
On Jun 27, 1998 this sequence version replaced gi:1261921.

Notes:
Details of M. tuberculosis sequencing at the Sanger Centre are
available on the World Wide Web.
(URL: http://www.sanger.ac.uk/Projects/M_tuberculosis/) CDS have
been renumbered from the original cosmid submissions but the old
gene designations are in brackets after the new gene numbers.
Gene prediction was based on a Hidden Markov Model of TB genes

Implemented in TParse (Krogh) supplemented with visual inspection
of positional base preference in codons, especially where there is
an increase in the observed/expected third position G + C.
CAUTION: In some cases we may not have predicted the correct
initiation codon. Where possible we choose an initiation codon
(atg, gtg, or ttg) which is preceded by an upstream ribosome
binding site sequence (optimally 5-13bp before the initiation
codon). If this cannot be identified we choose the most upstream
initiation codon.

FEATURES

source	location/Qualifiers
1..38110	/organism="Mycobacterium tuberculosis"
/strain="H37Rv"	/db_xref="taxon:1773"
<1..5460	/organism="Mycobacterium tuberculosis"
/strain="H37Rv"	/db_xref="taxon:1773"
source	
gene	
163..1599	/clone="Y190"
163..1599	/gene="glnA1"
163..1599	/note="RV2220, (MTCY190.31-MTCY427.01), len: 478, glnA1, probable glutamine synthetase, glnA, similar to many eg SW:GlnA_STRCO P15106 (71.4% identity in 475 aa overlap); contains PS00180 Glutamine synthetase signature 1, PS00181 Glutamine synthetase putative ATP-binding region signature, PS00182 Glutamine synthetase class-I adenylation site"
/codon_start=1	/transl_table=11
/product="glnA1"	/protein_id="CAA94665.1"
/db_xref="GI:3261568"	/translation="MTEKTPDDVFKLAKDEKVEVDVRFCDLPGIMQHFTIPASARDK SVFDDGLAFDGGISIRGFOSIHESDMLLPPEFARIDPFPAKTNLNFFVHPFPLE PYRDPENIRAKAEVYISGIDVAFVGAEEFYFDSPSRANGSFEVDALIE WMNMGATENDSGSPNNGYKRYRHNGGFPVAPNQQYDLDLRKMTNLNLSGFIIEKHH EVSGSGDAEINTQFNSLHAADMQLKTYIKNTAWONGKTYTFMRPLFGDGSCH CHOSLKGDAPIVDETGYAGLSDTAHYIGGLHHAHPSLAFNTPVNSYKRLVGH EAPINLVYSQRNSACVRIPIITSNPKAKLEFRSPSSGNPYLAESAMLMAGLDGKIK NKLEPQAPVKKDDELPEEASAIPOPPQLSDVIDRLEADHEYLEGGVFTNDLLET WISFKRENELEPNIRHPREFLYYDV"
382..>38110	/organism="Mycobacterium tuberculosis"
/strain="H37Rv"	/db_xref="taxon:1773"
955..1002	/clone="Y427"
955..1002	/gene="glnA1"
/note="PS00181 Glutamine synthetase putative ATP-binding region signature"	1342..1380
1342..1380	/gene="glnA1"
/note="PS00182 Glutamine synthetase class-I adenylation site"	complement(1917..4901)
complement(1917..4901)	/gene="glnE"
/gene="glnE"	complement(1917..4901)
/note="RV2221c, (MTCY190.32c-MTCY427.02c), len: 994, glnE, Probable glnE, similar to SW:GlnE_ECOLI P30870 glutamate-aminonitrogenase adenylyltransferase (24.4% identity in 721 aa overlap) and SW:GlnE_HAEIN P44419 (28.1% identity in 199 aa overlap); initiation codon uncertain unless at 39346 or 39247"	/codon_start=1
/transl_table=11	/product="glnE"
/product="glnE"	/protein_id="CAA94664.1"
/db_xref="GI:1261923"	

[illegible][illegible]

[illegible][illegible]

Db 591 CCCGGGCGTGGCCCGGAGTGAACCG--TCGAGTGCCTGGAGCGGCGGAGATGAGC 647
 QY 491 cgtacccgagcagatcgatgagcagctacagctccgagcgagctacccagctgaacagg 550
 Db 648 GGTACACGGCCAGCAGCTACCCCGGACAGCGGGCGAGACGAGACGAGCTGCTCGACG 707
 QY 551 tctaccgagcttgccgcaagagctgtgtacagcagctgagcttcaagcttcttgccaata 610
 Db 708 CTTACAGAGAGTTTCGGGAGGGCTGCG--GGGGGAGAGCGCCGAGAGCTGCTGGGCC-ACG 764
 QY 611 tcggtaccgctccgttcgacagagacatgacatggttcgccaagcgltaggtgaacgc 670
 Db 765 TCTCCACGGTTCGAGCGCGCATGAGACGCTCTCGCGGCTGCTGGGCGAGAGA 824
 QY 671 agatcaactacccctcgtgataagctacagcagcaagcttgagcagcagcttactgtgaagct 730
 Db 825 AGCTGACCTACGTGGGAGCGCTGTGAGGACCTTCTCTGGGCGGACCTACGCGCGTGTGT 884
 QY 731 tcggtactcatgtgagggcgatgtctctcagagcgagctacatcagcagcttagccaa 790
 Db 885 TCCCCGACGGAGCGGGCGGCTGTCTTGAGAGCGGCGGATGAGACCTCTGCTGGGCC 944
 QY 791 tcgaggaagaatcaagcaaatgagcgatctcagacgcttcaatgactacgagcgcg 850
 Db 945 GCGGCGTGAACCTTGAGACAGAGGAGGCTTCGAGACGCGTTCCACTTCTCGCGAAG 1004
 QY 851 acgagcgccgctcgccgagctgacctc-----ggagcagcagctcgccagctgag 901
 Db 1005 ACTGCGTGAAGCAGACCGGACTGCGCTCGGCGCAAGAGACACACCCCGGACAGGTGC 1064
 QY 902 tcaacgctacccagcagctgtgtgagcagctgtgtgagagcgagcgagtaagagctgagctc 961
 Db 1065 GCAAGAACCTCAAGTCTCTTTCGACAGACCTTGAGCGGAAAGCCCTCCCGCGGCGACG 1124
 QY 962 caagctgac-----ctgagctacgagcgagcgagcagcgagcagcagcagctgata 1015
 Db 1125 CCGAGCGCGCCAGCTCAGCGAATCCTTCGACACACCGGCGTGTGATGCGCGCATGTAGC 1184
 QY 1016 ggcctcagcgctggaagtaacctgaaccagctgtctgtgagcgctgagcgagcagcg 1075
 Db 1185 ACAGAGGCGCTGCGACACCTGCGGAGTCCCTCACCTCGGCGATCAAGAGAGAGACG 1244
 QY 1076 cggagcagctgt 1135
 Db 1245 GTGGGGCTGTGTATCTTTCGACAGCTACTAGAGCGGAGCGGCGGCGGCTGACA 1304
 QY 1136 gcaacgacagcagcagctgtcaacgagctccggtctgtgtgtgtgtgtgtgtgtgtgt 1195
 Db 1305 GCAACCTGATGTTCGCAAGCGCGCGCTGAACTGCTGACCTCCCGCGCTTCTGCT 1364
 QY 1196 cagcgagcctgt 1255
 Db 1365 CCGCGGAGAGTGTGCGGAGAGCGCTCCCGACCTTCGAGAGGAGGCTCCCGGCTTCCGCG 1424
 QY 1256 ggcagctacacagatccgcccccgagctgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1315
 Db 1425 AGGAGCTGTGTGTCTCTCTTAAC-----TGTGCTGTACTGTGCTGTGAAGCCCGCGG 1478
 QY 1316 cgaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1375
 Db 1479 GGGAGCGGCAACCGCATGAGAGCGGCGCGGCGCAACCGCATGCTGTGTGTGTGTGTGT 1538
 QY 1376 acgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1432
 Db 1539 GCGACCTCGGCGACCCCTTACCGTGTGGCGCAGAGCGCTCTCGACACCTACCTCCGCGC 1598
 QY 1433 cgcgtac 1492
 Db 1599 ACCTCTCTACCTACGAG 1658
 QY 1493 actctgagctgagctacatcttctcagagcagcagcagcagcagcagcagcagcagcagc 1552
 Db 1659 ACTTCGGGATCAACACTACTCTGCTGACGCGGACCGCGGAGAGAGAGAGAGAGAGAG 1718

QY 1553 cg 1554
 Db 1719 CG 1720
 RESULT 5
 LOCUS AR027260
 DEFINITION Sequence 7 from patent US 5856166.
 ACCESSION AR027260
 VERSION AR027260.1 GI:5938100
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE 1 (bases 1 to 1821)
 AUTHORS Barteld,D., Butler,M.J., Hadary,D., Jenish,D.L., Krieger,J.J., Malek,L.T., Soostmeyer,G., Walczyk,E., Kryszman,P. and Garven,S.
 TITLE Streptomyces proteases and methods for improved secretion of recombinantly-expressed proteins
 JOURNAL Patent: US 5856166-A 7 05-JAN-1999;
 FEATURES
 source Location/Qualifiers
 1..1821
 BASE COUNT 283 a 715 c 602 g 221 t
 ORIGIN
 Query Match 17.0%; Score 265.2; DB 81; Length 1821;
 Best Local Similarity 52.5%; Pred. No. 6,5e-28;
 Matches 788; Conservative 0; Mismatches 678; Indels 36; Gaps 8;
 QY 77 cagcgccccctgtgagcgagcagctgaagacccgagcgagcagcagcagcagcagcagc 136
 Db 231 CCGGGGCAAGGCGGCGGCGGAGGCGGACAGGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 290
 QY 137 cgt 196
 Db 291 AGGCGACGCGCGCGGAGCTGTCCCGTACACAGGAGGAGGAGGAGGAGGAGGAGGAG 350
 QY 197 aaattcgaactgtgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 256
 Db 351 GCGTCCGGGCTTCCAGTGTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 410
 QY 257 ggaacacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 310
 Db 411 ACGGGACGTCGCGGCTGCGGTCGCGGCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 470
 QY 311 ggaacactgt 370
 Db 471 GCTGTGCTGT 430
 QY 371 tggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 430
 Db 531 ACGGGCGATCGGCTACCGCGGAGAGGTCGCGCGGAGTGTGTGTGTGTGTGTGTGTGTGT 530
 QY 431 cgaagaggttcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 490
 Db 591 CCGGGGCGTGGCGCGGAGTGAACCG--TCGAGTGTCTGTGTGTGTGTGTGTGTGTGTGT 647
 QY 491 cgtacccgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 550
 Db 648 GGTACACGGCCAGCAGCTACCCCGGACAGCGGGCGAGAGAGAGAGAGAGAGAGAGAGAG 707
 QY 551 tctaccgagcttgccgcaagagctgtgtacagcagctgagcttcaagcttcttgccaata 610
 Db 708 CTTACAGAGAGTTTCGGGAGGGCTGCG--GGGGGAGAGCGCCGAGAGCTGCTGGGCC-ACG 764
 QY 611 tcggtaccgctccgttcgacagagacatgacatggttcgccaagcgltaggtgaacgc 670
 Db 765 TCTCCACGGTTCGAGCGCGCATGAGACGCTCTCGCGGCTGCTGGGCGAGAGA 824

small overlap between neighbouring submissions.
Cosmid 23B6
FEATURES
source
Location/Qualifiers
1..38366
/organism="Streptomyces coelicolor A3(2)"
/strain="A3(2)"
/db_xref="taxon:100226"
/clone="cosmid 23B6"
misc_feature
1..101
/note="nominal overlap with Streptomyces coelicolor cosmid Stp8"
RBS
37..41
46..795
/gene="2SC3B6.01"
46..795
/gene="2SC3B6.01"
/note="2SC3B6.01, unknown, len: 249 aa"
/codon_start=1
/transl_table=11
/product="hypothetical protein 2SC3B6.01"
/protein_id="CAC01307.1"
/db_xref="GI:9714465"
/translation="MSYLTSTKAGVASTGVHNGLVGSLAPHLAPDEWAGLTCTGCTP
LHMVVLNADGGAHPDPRCLPAAGRLNAGIRVGHDTAYGASLGGISDARYL
DMVYDGFLLAHCPSDRALPEVRRVTLTRALPDADHTGLGTHPCGVANNDOL
VTFSGTMADYRMSQAAMWSADYPPERFCHVHGLRGHLEEARLRARMGATICTFD
RTDHGRADPMETMECYMDEIVSRVGTGSE"
RBS
850..853
859..2037
/gene="2SC3B6.02"
859..2037
/gene="2SC3B6.02"
/note="2SC3B6.02, possible sulfurylase, len: 392 aa;
similar to TR:AAF3130 (EMBL:AF196567) pseudomonas
stutzeri putative sulfurylase, 391 aa; fasta scores: opt:
1433 z-score: 1624.0 E(): 0; 56.2% identity in 395 aa
overlap and to SW:MOEB_ECOLI (EMBL:M21151) Escherichia
coli molybdopterin biosynthesis Moeb protein, 249 aa;
blastp scores: Score = 526 (185.2 bits), Expect = 1.1e-50,
P = 1.1e-5, Identities = 106/249 (42%), Positives =
156/249 (62%). Contains Pfam match to entry PF00581
ThiF_family, ThiF family and PF00581 Rhodanese,
Rhodanese-like domain"
/codon_start=1
/transl_table=11
/product="putative sulfurylase"
/protein_id="CAC01308.1"
/db_xref="GI:9714466"
/translation="MSLPPLVPPAELFVDEVRRYSRHLIIPVGMGOKRLKNKVL
AVAGGIGSPITMLYLAAGVGTGLIIEPEVDESNLROGVHSOADIGRPKESRDT
IOGINPINDVITHOERLEADNMDIFSOCDLVDCGNDRATRYLVNDKAVLANKRYV
GSITFRDQASVFSWSEHGPCYKCLTPEPPPGVPSACAGVGLVLCASIGISQNEA
IKLLOGIGLPLGRMLKIDALEMTYRQVVRKDVPCAGVGNPVTLELLIDYAFGCV
SEEAQAADSTITPRLKEMTIDGENTELIDVRENEFEIVSIGARLIPANEFMG
SALENLPDKKIVLNCKTGVRSAAEVLAIVKSAFSDAVHVGGVIGVWQIPEKPVY
"
misc_feature
973..1383
/gene="2SC3B6.02"
/note="Pfam match to entry PF00899 ThiF_family, ThiF
family, score 216.00, E-value 5.4e-61"
1726..2010
/gene="2SC3B6.02"
/note="Pfam match to entry PF00581 Rhodanese,
Rhodanese-like domain, score 51.70, E-value 1.6e-11"
complement(2109..3650)
/gene="2SC3B6.03c"
/gene="2SC3B6.03c"
complement(2109..3650)
/gene="2SC3B6.03c"
/note="2SC3B6.03c, peptidase, len: 513 aa; identical to
TR:Q54399 (EMBL:U42759) Streptomyces lividans proteinase
SlpE, 513 aa and highly similar to Streptomyces coelicolor
2SC3B6.04c, 516 aa; fasta scores: opt: 1544 z-score:
1384.8 E(): 0; 50.2% identity in 526 aa overlap. Contains

misc_feature
misc_feature
CDS

RBS
gene
CDS

misc_feature
misc_feature
misc_feature

Query Match 12.0%; Score 187.4; DB 73; Length 38366;
Best local Similarity 51.1%; Pred. No. 7,7e-18;
Matches 695; Conservative 0; Mismatches 621; Indels 45; Gaps 9;

[illegible]

DB	4087	-----TGCACGGAATTGGCCGGTCCGCGAGCCGCCACACACCGGAGCTCCACGCGCG	4034
QY	1342	ggggctggcaaggttcgtgtgtgtgtcaccaccaccaccaccaccaccgagccactcgtatcgtcc	1401
Db	4033	GGGGCGGCGCGGATCTCTGTGTGATGGCAACACCGCGAGACCCGCGAGCCTCTACGAGGGC	3974
QY	1402	g-----gggtgacctgtgcccgcacgctggcgccacgcgtgtaccttgcgaagc	1452
Db	3973	GCCGGAGAGATGTGTGACGCGCTCGGCAAGGGCGCTCGCGCTCGAATCATCTACAGAGGC	3914
QY	1453	accacaacacactggtgttctgatgtgcaacacagtggtgtgactctggtgtatgcacct	1512
Db	3913	CAGGGCGACGGGCTCTACGACGATAGAAACAGGTGCTGTGACGAGCGGCTTCACGCGTAC	3854
QY	1513	ttcttcgaagagacctgtgcgccgcgagctgcgtggttcgc	1553
Db	3853	CTGCTCGACGGAAGAGTACCGCGCGCCAAAGACGCTGTCTC	3813
RESULT	8		
LOCUS	SCD8A	41800 bp	DNA
DEFINITION	Streptomyces coelicolor cosmid DBA.		BCF
ACCESSION	AL160331		13-MAR-2000
VERSION	AL160331.1	GI:7242748	
KEYWORDS	AAA family ATPase; ABC transporter ATP-binding protein; cytoskeletal RNA synthetase; glutamyltransferase; hydrolytic protein; integral membrane protein; lipoprotein; phosphate transport system regulatory protein; proteinase; response regulator; secreted protein; sensor kinase protein; transcriptional factor regulator; tRNA/4RNA methyltransferase.		
SOURCE	Streptomyces coelicolor A3(2).		
ORGANISM	Streptomyces coelicolor A3(2)		
REFERENCE	1 (bases 1 to 41800)		
AUTHORS	Redenbach, M., Kleiser, H. M., Denapite, D., Eichner, A., Cullum, J., Kinashi, N. and Hopwood, D. A.		
TITLE	A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb Streptomyces coelicolor A3(2) chromosome		
JOURNAL	Mol. Microbiol. 21 (1), 77-96 (1996)		
MEDLINE	97000351		
REFERENCE	2 (bases 1 to 41800)		
AUTHORS	Seeger, K. J. and Harris, D.		
JOURNAL	Unpublished		
AUTHORS	3 (bases 1 to 41800)		
TITLE	Cerdeno, A. M., Parkhill, J., Barrell, B. G. and Rajandream, M. A.		
JOURNAL	Direct Submission		
COMMENT	Submitted (13-MAR-2000) Streptomyces coelicolor sequencing project, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by prof. David A. Hopwood, [3] John Innes Centre, Norwich Research Park, Colney, Norwich, Norfolk NR4 7UH, UK		
NOTES:	Streptomyces coelicolor sequencing at The Sanger Centre is funded by the BBSRC and Beowulf Genomics		
	Details of S. coelicolor sequencing at the Sanger Centre are available on the World Wide Web.		
	(URL: http://www.sanger.ac.uk/Projects/S_coelicolor/)		
	CDS are numbered using the following system eg SC7B7.01c. SC (S. coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary strand).		
	The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous.		
	The length in codons is given for each CDS.		
	Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database.		
	The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon.		
	Gene prediction is based on positional base preference in codons		
	Nucleic Acids Research, 22(22):4768-4778(1994) and the Frameplot		

Program of Bibb et al., Gene 30:157-66(1984) as implemented at <http://www.nih.go.jp/jun/cgi-bin/frameplot.pl>. CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gta, tta or (att)) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.

IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions.

Cosmid D8A.

FEATURES

source 1.41800 Location/Qualifiers

organism="Streptomyces coelicolor A3(2)"

strain="A3(2)"

db_xref="taxon:100226"

clone="cosmid D8A"

complement(1..417)

gene /gene="SCDB8.01c"

complement(1..319)

gene="SCDB8.01c"

note="nominal overlap with S. coelicolor cosmid 25tcd6"

complement(<1..417)

CDS /gene="SCDB8.01c", probable phosphate transport system regulatory protein (fragment), len=139 aa; similar to C-terminal region of TR:O53833 (EMBL:AL022004) Mycobacterium tuberculosis phosphate transport system regulatory protein, Phoy-2, 213 aa; fasta scores: opt: 383 z-score: 454.2 E(): 7.2e-18; 46.8% identity in 139 aa overlap and of SW:PHOU_ECOLI (EMBL:X02723) Escherichia coli phosphate transport system regulatory protein, Phou, 241 aa; fasta scores: opt: 213 z-score: 258.9 E(): 5.5e-07; 28.1% identity in 128 aa overlap"

codon_start=1

transl_table=11

product="putative phosphate transport system regulatory protein (fragment)"

protein_id="CAB77322.1"

db_xref="GI:7242749"

translation="MRDAYHEELDSIGGLVEMARLYGSAIGRATTALIDSRLKAAE VTEADOKYDNOHELEARRATITLAROPVATDLRIYVTSLRMSADLERSGDLAQHVAK LARKYPERAVPHDHAITILEMGQLAQRLMAKAAEYI"

complement(426..431)

RBS 629

RBS 632..1912

gene /gene="SCDB8.02"

632..1912

gene="SCDB8.02"

note="SCDB8.02, possible sensor kinase protein, len: 426 aa; similar to SW:SEX3_MYCTU (EMBL:Y13628) Mycobacterium tuberculosis sensor-like histidine kinase Senx3 (EC 2.7.3.-), 410 aa; fasta scores: opt: 874 z-score: 939.7 E(): 0; 43.6% identity in 406 aa overlap and to C-terminal region of SW:PHOR_ECOLI (EMBL:X04704) Escherichia coli phosphate regulation sensor protein Phor, 431 aa; fasta scores: opt: 521 z-score: 562.5 E(): 6.7e-24; 36.8% identity in 261 aa overlap; also similar to various S. coelicolor sensor proteins, e.g. SCH22A.18c, 370 aa; fasta scores: opt: 415 z-score: 388.1 E(): 3.9e-16; 31.1% identity in 383 aa overlap. Contains Pfam match to entry PF00512 signal, Histidine kinase and possible N-terminal region signal peptide sequence"

codon_start=1

transl_table=11

product="putative sensor kinase protein"

protein_id="CAB77323.1"

db_xref="GI:7242750"

translation="MDVNAVAANAAGVLTGTVAVLAFFMSREQRKPTSLHTD PVLPGVDTVLVLSRAVLDADAVKASSAAALGIVRGRLAVPEPMOMADTR RGEIROYELDPRGNGRGALAVSARVAPLGSRLVLLVEDLLEARRIEVRDVF

misc_feature

note="Pfam match to entry PF00512 signal, Histidine kinase, score 217.80, E-value 1.6e-61"

CDS

gene /gene="SCDB8.03"

note="SCDB8.03, probable response regulator, len: 223 aa; similar to TR:O86269 (EMBL:MJ001103) Lactococcus lactis Arca protein, 233 aa; fasta scores: opt: 700 z-score: 810.0 E(): 0; 49.3% identity in 223 aa overlap and to TR:Q56180 (EMBL:U38917) Synecchococcus sp. response regulator Phor, 234 aa; fasta scores: opt: 687 z-score: 795.1 E(): 0; 48.0% identity in 229 aa overlap, and similar to many S. coelicolor response regulators, e.g. SCH22A.19c, 256 aa; fasta scores: opt: 687 z-score: 696.9 E(): 2.5e-33; 49.1% identity in 224 aa overlap. Contains Pfam matches to entries PF00072 response_reg, Response regulator receiver domain and PF00486 trans_reg_C, transcriptional regulatory protein, C terminal"

codon_start=1

transl_table=11

product="putative response regulator"

protein_id="CAB77324.1"

db_xref="GI:7242751"

translation="MLVVEDESFSDALSVMRKGEFVAIATTPGDLDEFERNGAD LVILDLAMPGLPGTEVCHQRLGRGNSVPVIMYATKSELDVGIGDDVTPFPSS BELVATIRAVLRGRGEPREVAAPALEAPVRMDNRHYVVGGRKVDPLKFEPLLEM LIRNAGRVLTIRQLIDRWGADYGDGTLVHVAKRLAKIPEDPGAVLVTVGGLG YKEP"

1918..2244

gene /gene="SCDB8.03"

note="Pfam match to entry PF00072 response_reg, Response regulator receiver domain, score 130.30, E-value 3.4e-35"

1918..2589

gene /gene="SCDB8.03"

2353..2571

gene="SCDB8.03"

note="Pfam match to entry PF00486 trans_reg_C, transcriptional regulatory protein, C terminal, score 117.10, E-value 9.8e-34"

complement(2750..3475)

gene="SCDB8.04c"

complement(2750..3475)

gene="SCDB8.04c"

note="SCDB8.04c, possible lipoprotein, len: 241 aa. Contains correctly situated match to Prosite entry PS00013 Prokaryotic membrane lipoprotein lipid attachment site and possible N-terminal region signal peptide sequence"

codon_start=1

transl_table=11

product="putative lipoprotein"

protein_id="CAB77325.1"

db_xref="GI:7242752"

translation="MSSSLRGLAALAAIAFASIASLACGAGHNAOTLEIKPDNAATS VGDIKVNAAVITPDDESTGPAVVASATLFPSCGKDPDLESTVAGTGGSLKIPAKG EKDGLTVAGGSLVIGCKDNAAAVLPSSREAVODGNMAKRTFNPSKTAVALRAVVP AESSFAEWGPESTPAAPGASAPSGEASATLSGEPADGASGTPADGRTAPGTDATG GATGDTGTGTGTGATGATGAGH"

complement(3401..3433)

gene="SCDB8.04c"

note="PS00013 Prokaryotic membrane lipoprotein lipid attachment site"

4177..4659

gene /gene="SCDB8.05"

4177..4659

gene="SCDB8.05"

note="SCDB8.05, probable transcriptional factor"

Db 706 TCCACGGCGGACTCCCGCGGAGAGACGAGCTTACCTGAAGACGCCGACTTCGGC 765
QY 568 caggaactgtgtacacggatagggttcaacttcttgccaatctcgttaccggttcgtc 627
Db 766 AGGGGCTGCAGAA---GGCCGGCGGCAACCTATGGAACACGCTTCGACGAGAACG 822
QY 628 gcaacgtacatgacatgtgtcccaagcgtltggttgatcatcatcactccctcgc 687
Db 823 GCCCGGCACTGACCTGATGCGGACGTCCTGGGCGACAGAGATGCACTTTCGGC 882
QY 688 tacagctacggcacaagtttggcaccgcttacccttgaaagtgcttgcgttactgtg 747
Db 883 ATCTCTACGCGACCGAAGCTGGCGGCTCTACGCCCATCTGTTCCCGGACGATGGCG 942
QY 748 gcatgtgtctctgacggtctatcgtaccagccgtttagcccaatcgagaacatcagc 807
Db 943 CGCGTATCTTCACGCGGGTGTGGACCCGGGCGCCACACGATGGGCCACGCCGAGAAC 1002
QY 808 caaatggcggatcttcaagacgcttcaatgactacgcgcgacttgcgcgctgcgcg 867
Db 1003 CAGGCCAGGGGGTTTCCACGCCGCGCTGGACGACTACTGTGACGACCGCCAGGAACC 1062
QY 868 gctcgcctcttgycacacgactcgcgcagtggttcaaccgttaccacgccttggttac 927
Db 1063 GACACAGGGGTCGGGAAAGATCGCGGCTG-----CTGGAAGCGG 1101
QY 928 cgcgtgtgtcagaagcgcggatgaagcgttcagatccagcgttggtcgttgctacgcgcg 987
Db 1102 CTGAGAGCGCCAGCCACTGCGCCACGCTCTCGCGGGGGAGGCTGACGACGACCTTCGCG 1161
QY 988 acgacgggacccaatacagcgtctgtacagccctcagcgcgtggaagtaactcgaatgtgt 1047
Db 1162 TTCACCGGCACTGCTGCTGCCGCTGTACAGGAGAGCGGCGCGGCCCTGACCAAGTGGC 1221
QY 1048 ctgtctgtgtcagc 1107
Db 1222 CTGAAAGCGGCGGCGGAGAG---GCGAGCGCTCGGAGTGTGGCCCTCGCGCGACGCGTAC 1278
QY 1108 gacgcgcggagatgacagcgggacatacagacacgaacgaacgcgttcaacgcgttcgcg 1167
Db 1279 AACGACGCTGATCCCTCGGGGGCGCTACGCGACGACGACCCACTCGCAAAAGGTCATATCG 1338
QY 1168 tgggtctatgtgcacacacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1227
Db 1339 TGCCTTGACACAAAGACAGAGGCGGACCGTGGAGAGACGAAGACTGCTGCGAGGTTTC 1398
QY 1228 cgtcaggtcgc 1287
Db 1399 GACAAAGTCTCTCCGCTTCGCGCGC-----TTCCTCGGCTGGGACACGCGCGGCTGG 1452
QY 1288 tgcgcgcgtgtgcgcgcgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1347
Db 1453 TGCACGAGACTGGCGGCTGGCGGTCAGACAGAAACCGCGAGGTGAGCGCGCGACGCGC 1512
QY 1348 ggaagaagtgcgt 1407
Db 1513 GCCCGGCACTGCTGCTGGTGGCAACGAGGCGACCGCGCCACGCGCTTACGAGGGCGCGCC 1572
QY 1408 gacctggccgc 1458
Db 1573 AGGATGGCGGACGACGCTGGGCAAGACGTCGCGCTGCTGCTGAGCGGCGGAGGGA 1632
QY 1459 cagactgcgt 1518
Db 1633 CAGCGGTGCTTACGAGGAAAGGAAAGCACTGCTGACTTCGCGGTGAGACGCGCTTACGTTG 1692
QY 1519 gacgggacactgcgc 1533
Db 1693 AAGGGGAGGCTGCGC 1707

RESULT 10
138961

LOCUS 138961 1777 bp DNA PAT 13-MAY-1997
DEFINITION Sequence 5 from patent US 5616485.
ACCESSION I38961
VERSION I38961.1 GI:2083441
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1777)
AUTHORS Hadary,D., Bartfield,D., Butler,M.J., Jenish,D., Krieger,T.,
Malek,I.T., Soostmeyer,G., and Walczyk,E.
TITLE Streptomyces proteases and improved streptomyces strains for
expression of peptides and polypeptides
JOURNAL Patent: US 5616485-A 5 01-APR-1997,
FEATURES
source location/Qualifiers
1..1777
BASE COUNT 288 a 607 c 666 g 216 t
ORIGIN
Query Match 10.1%; Score 158.2; DB 81; Length 1777;
Best Local Similarity 49.7%; Pred. No. 3.4e-13;
Matches 663; Conservative 0; Mismatches 618; Indels 54; Gaps 8;
QY 211 cgtctgcgcgcgcgt 270
Db 415 CGGTGGCGCCAGCGCTGAAAGGACCGCTGACCTGCTGCGACCCGACGAGCATGAT 474
QY 271 ttggcgt 327
Db 475 CTGCGCTGATCCGCTGCTCGCGGCGGAGGAGACGACCGCATCGCTGCTGCTGCTGCTGCT 534
QY 328 ccttggggacccgggc 387
Db 535 TTGGGGGCGCGCGCGCGCTCGCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 588
QY 388 gacacgcacatcttcgcgcacacttcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 447
Db 589 GTCTCTCTCTCTGACGACGCGCTGACGACCTGCTGACGCTGCTGACGCTGCTGCTGCTGCT 646
QY 448 tgcacccctcgttgc 507
Db 647 -CCAGGAGGCGCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 705
QY 508 atgacgcacacttcgc 567
Db 706 TCCACGGCGGACCTCCCGCGGAGAGACGAGGCTTACCTGAAGACGCGCGGACTTCGGC 765
QY 568 caggaactgtgtacacggatagggttcaacttcttgccaatctcgttaccggttcgtc 627
Db 766 AGGGGCTGCAGAA---GGCCGGCGGCAACCTATGGAACACGCTTCGACGAGAACG 822
QY 628 gcaacgtacatgacatgtgtcccaagcgtltggttgatcatcatcactccctcgc 687
Db 823 GCCCGGCACTGACCTGATGCGGACGTCCTGGGCGACAGAGATGCACTTTCGGC 882
QY 688 tacagctacggcacaagtttggcaccgcttacccttgaaagtgcttgcgttactgtg 747
Db 883 ATCTCTACGCGACCGAAGCTGGCGGCTCTACGCCCATCTGTTCCCGGACGATGGCG 942
QY 748 gcatgtgtctctgacggtctatcgtaccagccgtttagcccaatcgagaacatcagc 807
Db 943 CGCGTATCTTCACGCGGGTGTGGACCCGGGCGCCACACGATGGGCCACGCCGAGAAC 1002
QY 808 caaatggcggatcttcaagacgcttcaatgactacgcgcgacttgcgcgctgcgcg 867
Db 1003 CAGGCCAGGGGGTTTCCACGCCGCGCTGGACGACTTGTGATGACCGCGCGAGAACCC 1062
QY 868 gctcgcctcttgycacacgactcgcgcagtggttcaaccgttaccacgccttggttac 927
Db 1063 GACACAGGGGTCGGGAAAGATCGCGGCTG-----CTGGAAGCGG 1101


```

Db      883  ATCTCTACGAGCAGCACTCGGCGGGCGTCTACAGCCCATCTGTTCCCGGAGCAGTGGG  942
Qy      748  gcgatggtccctcgcagcggtatcgtaccagcggttaagcccaatcaggaagaatcagc  807
Db      943  CGCGTATCTCGACCGGGTGGGACCGGGGCGCCGACACGATGGGCCACGCGCGAAGC  1002
Qy      808  caaatgagggatcttcagagccgttcaatgactagccgagcagcagccggtcgcg  867
Db      1003  CAGGCCAGGGGTTTCCAGCGCGGCTGACGACTCTGAGTGCAGCCGCGCAGAAACC  1062
Qy      868  gctcgccctctggcagcagcagcagcagtggttaacgcgtaccagccctgtgtgac  927
Db      1063  GAACAGGGGTCGCCGGAAGATCCGCCGGCTG-----CTGAGACGG  1101
Qy      928  ccgctgtgtcagaagcgcggtaagacgttcgtaccagcttgagcgtacgcagcagc  987
Db      1102  CTGAGACGGCGACCACTGCCCCAGTCTCGCGGGGCGGAGGTACGACAGACCTCGCG  1161
Qy      988  acgacgagggacatcaacagcgtgtcagacccctcagcgtctgaagttaactgacag  1047
Db      1162  TTCACCGGCTGCTGCTGCTGTAACAGCAGAGCGGCTGGCGGCCCTGACCACTGCG  1221
Qy      1048  ctgctgtgagctgaagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc  1107
Db      1222  CTGAGAGCGCGCGAGAGG---CGACGCGCTCGGACTTCTGCGCCCTCGCGACGCTAC  1278
Qy      1108  gacgcgcggatgacagcagcggaactaacagacagacagcagcagcgttcaacgcgc  1167
Db      1279  AACGACCGTATCTCCCTCGGGCGCTACGCGACGACGACCCACTCGCAAGGCTCATATG  1338
Qy      1168  tgcgtctatgcccacacacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc  1227
Db      1339  TTCCTGTGAGACAGACAGAGCGCCGACCGTGTGAGAGACAGAACGCTGCTGCGAGCTT  1398
Qy      1228  cgtcaggtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc  1287
Db      1399  GAGAGAGTCTTCCTCTCTTCGGCGC-----TTCTCTGCGCTGGGACACGGCGGGTGG  1452
Qy      1288  tgcgcgcgtgtgacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc  1347
Db      1453  TTCACGACGATCGCGCGCTGCGCTGACGACGACGCGGAGTGTGAGCGCGCGCGACG  1512
Qy      1348  ggcgaagtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt  1407
Db      1513  GCCCGGCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT  1572
Qy      1408  gacctgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc  1458
Db      1573  AGGATGTGCGAGGAGCTGGGCAAGGACGTGCGGTGTGTGTGTGTGTGTGTGTGTGTGT  1632
Qy      1459  cactcgtgcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt  1518
Db      1633  CACGGTGCTACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG  1692
Qy      1519  gacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc  1573
Db      1693  AAGGGACGCTGCGC  1707

```

```

RESULT 12
LOCUS   SC5H1 36583 bp DNA BCT 06-JUN-2000
DEFINITION Streptomyces coelicolor cosmid 5H1.
ACCESSION AL049863
VERSION AL049863.2 GI:8347022
KEYWORDS add; adenosine deaminase; AMP-dependent Clp protease; bldA codon;
          chf; chitinase; clpP3; clpP4; DNA-binding protein; hydroxylase;
          ion channel subunit; isomerase; myo-inositol dehydrogenase;
          oxidoreductase; protease; ribonuclease H; RNA polymerase sigma
          factor; rnhA; serine/threonine protein kinase; transcriptional
          regulator.
SOURCE  Streptomyces coelicolor A3(2).
ORGANISM Streptomyces coelicolor A3(2)

```

REFERENCE

Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces. 1 (bases 1 to 36583) Redenbach, M., Kieser, H.M., Denaplatte, D., Elchner, A., Cullum, J., Kinashi, H. and Hopwood, D.A. A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb Streptomyces coelicolor A3(2) chromosome Mol. Microbiol. 21 (1), 77-96 (1996)

JOURNAL

Medline 2 (bases 1 to 36583) Oliver, K. and Harris, D. Unpublished 3 (bases 1 to 36583) James, K.D., Parkhill, J., Barrell, B.G. and Rajandream, M.A. Direct Submission Submitted (10-MAY-1999) Streptomyces coelicolor sequencing project, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof. David A. Hopwood, [3] John Innes Centre, Norwich Research Park, Colney, Norwich, Norfolk NR4 7UH, UK On Jun 8, 2000 this sequence version replaced g1:4835303.

COMMENT

Notes: Streptomyces coelicolor sequencing at The Sanger Centre is funded by the BBSRC. Details of S. coelicolor sequencing at the Sanger Centre are available on the World Wide Web. (URL: <http://www.sanger.ac.uk/Projects/S.coelicolor/>)

CDS are numbered using the following system eg SC7B7.0lc. SC (S. coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary strand). The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous. The length in codons is given for each CDS. Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database. The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon. Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(22):4768-4778(1994)) and the Frameplot program of Blab et al., Gene 30:157-66(1984) as implemented at <http://www.nih.go.jp/jun/cgi-bin/frameplot.pl>. CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, ttg or (att)) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon. IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions. Cosmid 5H1 lies between 262 and 588 on the AseI-A genomic restriction fragment.

FEATURES

source 1. 36583 /organism="Streptomyces coelicolor A3(2)" /strain="A3(2)" /db_xref="taxon:100226" /clone="cosmid 5H1" complement(179..238) /note="55 bp imperfect inverted repeat" complement(209..1246) /gene="SC5H1.41" complement(1209..1246) /gene="SC5H1.41" /note="SC5H1.41, hypothetical protein, len: 345 aa; unknown function, similar to part of SCE29.03 (EMBL:AL035707) S.coelicolor putative phosphotransferase (300 aa), fasta scores: opt: 433 z-score: 437.4 E(1): 4.9e-17, 37.2% identity in 269 aa overlap. Contains rna (leucine) codon, possible target for bldA regulation. There are imperfect inverted repeats at each end of the

repeat_unit

gene repeat_unit CDS

```
CDS
/codon_start=1
/label_table=11
/label="SC5H1.41"
/product="hypothetical protein"
/protein_id="CAB42966.1"
/db_xref="GI:4835344"
/db_xref="SPTREMBL:Q9X7U8"
/translation="MLGAGRARLGAGPGTGPLSGRIDLSGPGCAOLRTAATAVAHRICP
EFSVYVLRSGRSVLLVGTGTSTAVAKCLVDHSPAMERSRHEIGASRSPVROPP
VRYPRIADPDRNGSTLIERMPGVVALOHNPAPRADIRAIAGVCRHMPAPSA
GTEDAPLDVARIASRYHEIGLLDROLDGLOKLLHGTAHAGRGOMGCPGCMALLS
VMSAPGVYVDMEHAGRTPGYDLATLMAVLGDAPVARKOISIOASTGTARDNFL
VNLMVLVTRIRITYETAVQSRMHDVGAATGAPGAAPTEGEORLLRLRHDDCOLA
RAVRAVAVGT"
complement(1022..1024)
/misc_feature
/gene="SC5H1.41"
/feature="TTR (leucine) codon, possible target for bida
regulation"
complement(1287..1346)
/feature="55 bp imperfect inverted repeat"
1648..3156
/gene="SC5H1.40c"
1648..3156
/gene="SC5H1.40c"
/feature="SC5H1.40c"
/feature="SC5H1.40c, possible regulatory protein, len: 502
aa, similar to e.g. SW:SP15_STRGR (EMBL:M32687)
Streptomyces griseus sporulation protein (529 aa), fasta
scores: opt: 394 z-score: 452.2 E(): 7.2e-18, 27.9%
identity in 530 aa. Similar to SC7A1.26 (EMBL:AL034447)
S.coelicolor possible transcriptional regulator (500 aa)
(43.0% identity in 491 aa overlap)"
/codon_start=1
/label_table=11
/label="SC5H1.40c"
/product="putative regulatory protein"
/protein_id="CAB42965.1"
/db_xref="GI:4835343"
/db_xref="SPTREMBL:Q9X7U7"
/translation="MSKEPPTRLSLDGLAGMSKGLARLVNRQAAMGHPOLATPTS
RVRNMDGEIPRDPYPRVLAALFTERLGRVYTEDLGLVRHGRACKRRPDSEEPD
GVWAPERTRAVLLEFTGMDLNRKGLVAGALAGSALSSAMHDWLHTDALAD
APDLQPLHADPAFGDFRYEAPIGSQEVELESEVEFRAMDAAKRGGLORAKVVOQL
NEVGMILAYHNHPHILQRLMGVAAANLAVLGMNSHVGLEPTQKQFVLAHAARAGG
DRPRAGEALSRAARQVHLKQPEALDKMLAAGSGGEVLLPTKMLVLTIEAQAAS
MCKGOMARTRIGRAEDLVSDKADVPDPMPOMOFKREDLYGMQALVRLTAEERPA
AAHQHAYADKALALRIDGRORSKTFDYLSMASACFINDDPQADRYARLALVSGSNS
QRTWRLQRMTRILTAETAGTPIQELREELIKLPLKGSFRGKPGGTVAFA"
complement(3176..4492)
/gene="SC5H1.39"
/feature="SC5H1.39"
/feature="SC5H1.39"
/feature="SC5H1.39, hypothetical protein, len: 438 aa;
unknown function, probable CDS suggested by positional
base preference, GC frame analysis and amino acid
composition"
/codon_start=1
/label_table=11
/label="SC5H1.39"
/product="hypothetical protein"
/protein_id="CAB42964.1"
/db_xref="GI:4835342"
/db_xref="SPTREMBL:Q9X7U6"
/translation="MSSHLSDHPAASOGGSVEALISOTRRLRGDVAVRNSODD
TTPDRGRORALYDLALHOLITDLDALHOLRSPAPRAGSILSVGSAENMLLDE
AKMSALFHLILGSDPASVPLDELPSLVTEDRPKLTAMVTDCLVDGSRIDGEEFVY
RDLGSRVRYHMMGEPEVLADGSPASMAVLRDVGELRSQRAVSETRDLRRRPAW
HGLLREPDGGRGLADGAHLPTSATGRSRSGVADCPGLGADDETLITVGDGFLP
EADYGPVAVLIGRLGLALGTRPRRLRLAEILARLAPVADGVVGVGYPDRHTL
TAAHAGSPAPLIFRGSTHALSGPGRAPADPGSCGAEATLTAAGDILLHTALPEEA
VEHLTLARLTRAADSGVRLVTRLELDGFERODACVLIARVTP"
complement(4765..5793)
/gene="SC5H1.38"
gene
```

```
CDS
complement(4765..5793)
/gene="SC5H1.38"
/feature="SC5H1.38"
/feature="SC5H1.38, probable myo-inositol dehydrogenase, len:
342 aa; similar to many e.g. SW:M12D_BACSU (EMBL:M76431),
14d, Bacillus subtilis myo-inositol 2-dehydrogenase (344
aa), fasta scores: opt: 864 z-score: 958.5 E(): 0, 38.18
identity in 336 aa overlap. Also similar to SW:STR1_STRGR
(EMBL:Y00459), str1, Streptomyces griseus streptomycin
biosynthesis myo-inositol oxidoreductase (348 aa) (32.7%
identity in 343 aa overlap). Weakly similar to TR:069945
(EMBL:AI023862) S.coelicolor possible oxidoreductase (430
aa) (29.8% identity in 228 aa overlap). Contains Pfam
match to entry PF01408 Gfo_IDH_MocA, Oxidoreductase
family, score 153.50, E-value 3.6e-42"
/codon_start=1
/label_table=11
/label="SC5H1.38"
/product="putative myo-inositol dehydrogenase"
/protein_id="CAB42963.1"
/db_xref="GI:4835341"
/db_xref="SPTREMBL:Q9X7U5"
/translation="MSELIGVAVLGVAGHMGADHRRVDQVSGARVAVAADPAERAK
EANGTIGTGRTVTHPVDFAALDAPVEVAULIASPREAHEBALAFAPGLPYLCKP
MAPNSGALRVVEAEARLGRRLAQIGFMRRYDAETQLSLDGRGLRPLHLCVHR
MVSSPHTFSAMLINSSVSHEDAAKMLLGOELASVTVLRPRPSAGAPGLDLPOLV
FETEGGAVVDVEVFVNCGFCGVEACVACAGARIGAAHTMVTAAAGRAEVPDQY
LVFEADAVREVQSWDATRRLGLVTGPTMGVAAVAAGVAGALDGTVRTVPDMP
RSLHDRA"
complement(5062..5781)
/gene="SC5H1.38"
/feature="Pfam match to entry PF01408 Gfo_IDH_MocA,
Oxidoreductase family, score 153.50, E-value 3.6e-42"
complement(5804..5808)
/feature="possible RBS"
complement(5917..7011)
/gene="SC5H1.37"
complement(5917..7011)
/gene="SC5H1.37"
/feature="SC5H1.37"
/feature="SC5H1.37"
/feature="SC5H1.37, hypothetical protein, len: 364 aa;
unknown function, similar to TR:P73120 (EMBL:D90903)
Synchocystis sp. hypothetical protein (358 aa), fasta
scores: opt: 492 z-score: 559.6 E(): 7.6e-24, 31.6%
Query Match 8.6%; Score 134.6; DB 73; Length 36583;
Best Local Similarity 49.0%; Pred. No. 1.4e-10;
Matches 430; Conservative 0; Mismatches 429; Indels 18; Gaps 2;
QY 214 tgcgcagcagtgctcgtcccgctgactacgacaaaccggtgagacacaagcaagtgtg 273
||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 8511 TCGGGCAGAGGTACACCGTCCCTCGATTACTCGCGCCCGCGGAGAACCTCGACGTG 8452
QY 274 gcggtgtagtcggtcccgagcgagcagcagattcggagcactgtcgtgcaatcccggtg 333
||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 8451 GCCCTGGCCCGGTACCGGGCGGCGGACTCGCGGGGCTCGGTCTCGTGAACCTTCGGC 8392
QY 334 ggaaccggggcgctcgcgcggtcgacatgctgcgcgcta tggcacccgagatcgagcaacc 393
||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 8391 GGCCCGCGCGGCTCGGCAATCAACGAACCTGCGCTGGCGCGCAAGAGTTCATGCACCTC 8332
QY 394 gacattctcgacactcgacacttggtagcttcgacccgagagggatcggtcgacactgacc 453
||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 8331 ACCAA-----CGGCTACGACGCTGTGACTTCGACCCCGCGCGCTCGCTCTCC 8278
QY 454 ccgctgttggtggtgtcgacacgagcgcgagtttgaagcgtaacggcgcgatcgagcaacc 513
||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 8277 CCGGTCTCTGCGGACCGGCGACCTGGAAGATATGAGG-----CGAGCGGAC 8230
QY 514 gactcaagtcgcgcgcggtcacccacgctcgaaacaggtctaccgcgcaagttggccggagc 573
||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 8229 GGGGACGGGGGATGACCGACCCCGAGGGCGTCTGGAACGACTGCGGACGCGGGGGCC 8170
QY 574 tgtgttacccgagtggtgctcagcttcgtgccaatatcgtaaccgcttcgagcagcg 633
||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```


Db 8169 GAGTGGCCCAAGTACTCCGGCCCGGTGTCGCCACATAGGACAGGTGCGAGCGGCCCGC 8110
 QY 634 gaacatgacatggttcgccaaagcgttagtgatgacatagatcaactaacctcgatagc 633
 Db 8109 GACATGACACGTGATGACCCCGGGCCCTTCGGCGACGACCGGCTCAACTGACCTGCTTCG 8050
 QY 694 taagagcccaagtgttgccacgcgttacttgaaagcgtgtgtaactatgtgaggcgatg 753
 Db 8049 TAGGGACCCGGTGGGGGGGCTTACGGCGCGGGTCCCGGACAGGTGCGCGGATG 7990
 QY 754 gtctcagcagcgctatcgtatccagccgttagcccaatcaggaagacatcagccaatg 813
 Db 7989 GTCTTGACGCGGTGGAGACGCTGACGGAACCGCTGGCGACAGGGGCGTGGCGGGCG 7930
 QY 814 gggggaattcagacgcgttcaatgactacgcgcgcgcgcgcgcgcgcgcgcgcgcgc 873
 Db 7929 CGGGGGGACGACGACCGGCTGAGAACTTCTGACGTGCGTGCAGAGACGTGCGCTGC 7870
 QY 874 cctctggagccacgcgttcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 933
 Db 7869 CCGTTGCGCCAGACACCGCCGGGACCGCCGGGACAGGTGACGGGCTGCTGCTCTC 7810
 QY 934 gtgcagaagcgcgcgttagaacgctgcacgtgcgcgcgcgcgcgcgcgcgcgcgcgcgc 993
 Db 7809 GACTCGGACCGGGTGGCGGTTCGGCGAGCCCTTACCGGGGACGACGACATGCTGGC 7750
 QY 994 ggcacatcaagcgcgtgtacagccctcagcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1053
 Db 7749 GCATGCGGCGAGGCGCTGTGACAGCCGGAGTGTGCGCCCTGCGTGGAGCGGACCTGCG 7690
 QY 1054 gggctcagc 1090
 Db 7689 CAGTTGCTCGAGAGACGGTGCACACCGCGGCTGAGAG 7653

RESULT 13
 LOCUS SC9C7 31360 bp DNA BCT 12-JAN-1999
 DEFINITION Streptomyces coelicolor cosmid 9C7.
 ACCESSION AL033161
 VERSION AL033161.1 GI:4154059
 KEYWORDS acetyltransferase; acyl-CoA dehydrogenase; acyl-peptide hydrolase; aminolipase; ccr; coenzyme B12-dependent mutase; crotonyl CoA reductase; efflux protein; lyase; meaa; oxidoreductase; peptidase; tetr family; transcriptional regulator.
 SOURCE Streptomyces coelicolor A3(2)
 ORGANISM Streptomyces coelicolor A3(2)
 Bacteria; Firmicutes; Actinobacteria; Actinomycetidae; Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
 1 (bases 1 to 31360)
 Seeger, K.J. and Harris, D.
 Unpublished
 2 (bases 1 to 31360)
 Bentley, S.D., Parkhill, J., Barrell, B.G. and Rajandream, M.A.
 Direct Submission
 Submitted (08-JAN-1999) Streptomyces coelicolor sequencing project, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: barrelles@sanger.ac.uk Cosmids supplied by Prof. David A. Hopwood, [3] John Innes Centre, Norwich Research Park, Colney, Norwich, Norfolk NR4 7UH, UK
 3 (bases 1 to 31360)
 Redenbach, M., Kiese, H.M., Denapate, D., Eichner, A., Cullum, J., Kinsahl, H. and Hopwood, D.A.
 A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb Streptomyces coelicolor A3(2) chromosome
 Mol. Microbiol. 21 (1), 77-96 (1996)
 97000351
 Notes:
 Streptomyces coelicolor sequencing at The Sanger Centre is funded by the BBSRC.
 Details of S. coelicolor sequencing at the Sanger Centre are available on the World Wide Web.
 (URL: http://www.sanger.ac.uk/Projects/S_coelicolor/) CDS are

FEATURES

numbered using the following system eg SC9C7.01c. SC (S. coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary strand).
 The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous. The length in codons is given for each CDS.
 Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database. The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon. Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(22):4768-4778(1994)) as implemented at <http://www.nhl.go.jp/jun/cgi-bin/frameplot.pl>. CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, ttg or (att)) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.
 IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions. Cosmid 9C7 lies between 1A11 and 1E6 on the AseI-A genomic restriction fragment.
 Location/Qualifiers
 1..31360
 /organism="Streptomyces coelicolor A3(2)"
 /strain="A3(2)"
 /db_xref="taxon:100226"
 /clone="cosmid 9C7"
 /complement(1..147)
 /gene="SC9C7.01c"
 /complement(<1..147)
 /gene="SC9C7.01c"
 /note="SC9C7.01c, partial CDS, unknown, len: 49aa; overlapping extreme N-terminal region of partial CDS, SC1A11.2c from Streptomyces coelicolor cosmid 1A11."
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein SC9C7.01c"
 /protein_id="CAA22713.1"
 /db_xref="GI:4154060"
 /translation="MTADYATYIAGLPVLAAGAAVFRDAGRVLLVPEPNREGWALPGTIT"
 1..112
 /note="Overlap with SC1A11 Streptomyces coelicolor cosmid 1A11."
 332..1474
 /gene="SC9C7.02"
 332..1474
 /note="SC9C7.02"
 /gene="SC9C7.02"
 /note="SC9C7.02, conserved hypothetical protein, len: 380aa; similar to many eg. SW:YXAA_BACSU hypothetical protein from Bacillus subtilis (382 aa) fasta scores: opt: 1045, z-score: 998.9, E(): 0, (46.3% identity in 374 aa overlap)."
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="CAA22714.1"
 /db_xref="GI:4154061"
 /translation="MDGTRRVLAADKFKGSLTAEVAVRTAGLRVVDVLEALP VADGGGTVAAVAAVAGFERREAVAGPLDEEYTAVALDGPAAVEMASGLRIPE GVLAPLTSVTGSGELLRALDAGARTIVFGGSGATGGGAMLAAGALDNGE PVAPGGGAGIAGVADISGLDSRLSIVELVLAADNDPLTPGKGAAYGKGAAPD DVTALDAIAHFAKYLERTEGYGARAETAAAPGGAAGGIFGAMULGARRPPIEV MLDVGFAPALERADLVITGESLDEQTLHGAPGAAVAAAAGAEVAVVAGRLALP ABVLGAGIRRAYVPLTDVDPVAKCIADAGPILTEVAESIAIDFLA"
 /complement(1744..2526)
 /gene="SC9C7.03c"
 /complement(1744..2526)
 CDS
 gene
 CDS

Db 15482 CTTGCGCTTCAACGACCGCCACCGGGGTTCTGCGCTGGATCGCCGAGACGACGCGGG 15423
 QY 870 ctgacctctggggaacgactcggccagtggtgtcaaccgctaccagccctgtgtaccc 929
 Db 15422 CTACGGTCTGGGAGGACGCCCGAGGGGTGAGGCGCTGTGCTACGCGCATGGGAGCGGG 15363
 QY 930 gctgtgacgagacgaggaagacgtcgtatcaacgttgcccttgatgacgagcagcag 989
 Db 15362 CTTGGCGCAAGCGGCGCGCGCGGGAAGC-----TCGGCGCTCCGAGCTGGGA 15315
 QY 990 gacgggacacatacaacgagctgtacagccctcagcgtgtgaagtacatgacagtggtct 1049
 Db 15314 GGAACACTTACATCCGCGGGGGGTACTACAAAGGCTACTGGCCCTACTCTGGCCGAGCGCTT 15255
 QY 1050 gctgtgagctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1109
 Db 15254 CGGCTCTACGCGCAAGGAGAGACACCGCGCGCTGTGAGGCGCTACGACAACTTCCG 15195
 QY 1110 cggcgggagatgcagaggggacactacagcaagcagcagcagcagcagcagcagcagcag 1169
 Db 15194 CGGCTTGAGCGCTCTCGGGGAGACAGGCTAC-----AGGCTCTACACCGCGGTGCACTG 15141
 QY 1170 cgtctatgcgccaacacacgagcagcagcagcagcagcagcagcagcagcagcagcagc 1229
 Db 15140 CGCGGACGCGCGCTGTGGCGGGGAGCTGGAGCAGTGGCGGGAGACAGCTGGGACCTGTA 15081
 QY 1230 tcaggtcgcgcgcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 1289
 Db 15080 CGGAGAGGACCTTCAT-----GGTGTGAGAACAGCGCTGTGACACCGCGCGTGG 15030
 QY 1290 cggcgtgtgagcgggtgctcggaacgtcgaagcggcagcggcggcggcggcggcggcggc 1349
 Db 15029 CGGCTTCTGGCCACCGGGTCCGGGACGCGGTGAGCTGCGCAAC---GGGAGCTGCC 14973
 QY 1350 caaggtcgtcgt 1409
 Db 14972 CCGGCTCTGCTGTTCAGAGCGACGCGCGAGCGCGGCTTACGAGGGCGCGCTCGG 14913
 QY 1410 cctgtgcgcgacagctgt 1463
 Db 14912 CGGCGACCGGCTGCTGCGGACTCCGCTCTGCTGCTGAGAGGCGCGGACACCGG 14853
 QY 1464 tgcggt 1523
 Db 14852 CATCAGCTGAGCGGAGGAGCTGCTTGAAGCAGACCTTGGCCGCTTACCTGACGCGG 14793
 QY 1524 gacctgtgcgc 1534
 Db 14792 AACGGTGGCG 14782

RESULT 14
 AR027257
 LOCUS AR027257 1908 bp DNA
 DEFINITION Sequence 1 from patent US 5856166.
 ACCESSION AR027257
 VERSION AR027257.1 GI:5938097
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE 1 (bases 1 to 1908)
 AUTHORS Bartfeld,D., Butler,M.J., Hadary,D., Jenish,D.L., Krieger,T.J.,
 Malek,L.T., Soostmeyer,G., Walczyk,E., Krysgman,P. and Garven,S.
 TITLE Striptomycos proteases and methods for improved secretion of
 recombinantly-expressed proteins
 JOURNAL Patent: US 5856166-A 1 05-JAN-1999;
 FEATURES
 source Location/Qualifiers
 1..1908
 BASE COUNT 322 a 753 c 608 g 225 t
 ORIGIN

Query Match 7.3%; Score 113.2; DB 81; Length 1908;
 Best Local Similarity 46.1%; Pred. No. 56-07;
 Matches 669; Conservative 0; Mismatches 743; Indels 40; Gaps 7;

QY 12 ggcggcagacgctgtgagctcgt 71
 Db 197 GCACTGTGTACCGCGCGCTGTGATCGCGGCGCTGTGCGCACCGCGCGGAGCGCGCC 256
 QY 72 cctagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 131
 Db 257 CCGGCGGAGGCGGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 316
 QY 132 tgcgcgctcgt 191
 Db 317 GCG 372
 QY 192 cagcgaatctgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 251
 Db 373 GAACCTGCGCAAGCGCGCTACCTGCGCTACCTGCGCTACCTGCGCTACCTGCGCTACCT 432
 QY 252 cgttggagacacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 302
 Db 433 GTACGGGACAGAGATCAGCTGCGCGCTGCGCGCTGCGCGCTGCGCGCTGCGCGCTGCG 492
 QY 303 gcatctcggagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 362
 Db 493 GCGCGAGGCGCGCTGATGTACAAACCGCGCGCTGCGCGCTGCGCGCTGCGCGCTGCG 552
 QY 363 cgcgcgtatgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 419
 Db 553 GCGCGCGCTGCGAGACAG 612
 QY 420 gggctcgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 479
 Db 613 CCGGCTTGCAGCG 672
 QY 480 cgaatcgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 539
 Db 673 GTTGTCAAGGACCGCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 732
 QY 540 cgtgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 599
 Db 733 GCGC---AAGCTGCGCGGAGTACCGCGGAGGCTCTTGCAGCGGAGCGGAGTGCCT 789
 QY 600 cttggccaatatcgt 659
 Db 790 CCGCGCATGAGCC---AGCGCGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 846
 QY 660 aggtgacgactcagatcaactcagatcaactcagatcaactcagatcaactcagatcaact 719
 Db 847 CGGCGGAGAGAGAGCTCACTACCTGCGCGCTCTCTACGCGACCTACCTGCGCGCTGTA 906
 QY 720 cctgtgaagcgt 779
 Db 907 CGGACCTCTTCCCGGAGCAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 966
 QY 780 cgttagcccaatcagaggaagcagcagcagcagcagcagcagcagcagcagcagcagc 839
 Db 967 CCGCGCAAGATCTGTGTACAGGCGCAAC-----CTGGACCAAGAGCTGCC 1011
 QY 840 ctaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 899
 Db 1012 CTTCGAGGCGCGCTGGAAGAGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1071
 QY 900 ggttcaacgctacacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 959
 Db 1072 CTTGCGGACACCGCGCGCGGAGGTCCAGGACCACTGTGAGAGCTGCGCGCGCGCGCG 1131
 QY 960 tcaagctgtgcgt 1019
 Db 1132 GAAGAAAGCGCGTGGGCGCGCTGCTGCGAGCGCGCGGAGCTGATCTCTTCTCAGAGCGC 1191

1020 tcagcgctggaagtaacctgaacagtggtctctggtggcgctgcagcgcaagcgcccg 1079
1192 cccgtactagacatcccgcttggtggcgacgacggagatcttcagacaaagtaacgtccggcg 1251
1080 cgaactctgtgtctcttcgcaagactatga---cgccggagatgcagagcgactacag 1136
1252 CGACACCAGCGCTGCTCGACGCGCGGACCGGACCTGTCCGACACCGCGGGCAACGCG 1311
1137 caagcgacaggaagcggttcaacgcggtctcgtgtgctatgcgcacacacgcgcgacttc 1196
1312 CTCCGCGGAGAAAGGCAACGCCGCTTACAGCGCGCTGAGTGCAACCGACGCAAGTGCGC 1371
1197 agcgacctgtgtgtgcgcgcgcgaacagatctcagatgcgcgcgcgcgttcttactagc 1256
1372 CGCAACTGGGCGACCTGGGACCGGGAACACCGCGGCTCCAGCGGACGACCGCTTCAT 1431
1257 gcaattacccggtatcgcccccgcgatctgtgcgcgtctgtgcgcgtgtgcgcgcgcgc 1316
1432 GACCTGGGCGCAACGCTGGATGAACCTGCCCTGTGCCACCGCGCGGTCAAGCAGCAGAC 1491
1317 gacgcgcaaccccg 1376
1492 CCCGCTAAGCTGAAGACCGGCAAGGACTTCCGCGGCTGATCTGTCAGTCCAGTCCGAGCG 1551
1377 cgaccgcgcacatcgcatacagtcgcggttagacttgcgcgcgcgcgcgcgcgcgcgcgc 1436
1552 TGACGCGCGGACCGCGGACGAGGCGCGCGTGAACCTGACACGCGGTCCGCGGATCCCG 1611
1437 gatcactctga 1448
1612 CTGTATCAGCA 1623

RESULT 15
LOCUS 138959 1908 bp DNA PAT 13-MAY-1997
DEFINITION Sequence 1 from patent US 5616485.
ACCESSION 138959
VERSION 138959.1 GI:2083439
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1908)
AUTHORS Hadley,D., Bartfeld,D., Butler,M.J., Jenish,D., Krieger,T.,
Malek,I.T., Soostmeyer,G. and Walczyk,E.
TITLE Streptomyces proteases and improved streptomyces strains for
expression of peptides and polypeptides
JOURNAL Patent: US 5616485-A 1 01-APR-1997;
NATURES Location/Qualifiers
1. 1908
Source /organism="unknown"
BASP COUNT 322 a 753 c 608 g 225 t
ORIGIN

Query Match 7.3%: Score 113.2; DB 81; Length 1908;
Best Local Similarity 46.1%: Pred. No. 5e-07;
Matches 669; Conservative 0; Mismatches 743; Indels 40; Gaps 7;

12 ggcgcgacagacgcttgagctgcgcgtctgtcttctcgtgtgtgtgtgtgtgtgtgtgtgtgt 71
197 GCACTGTGTCACCGCACGCGCATGATCGCGGCGCGCTCTCGGACCGCGGAGCGCGCGCGCG 256
72 ccttagcgagcgccccgctgtgc 131
257 CCGGCGCACGCGCACGCGGCAACGCGGAGCTGGGACCGGAGGCGCGGCTGCCGCGCATC 316
132 tgcgcgcgtgtgtgc 191
317 GCG 372
192 cagcgaaatcgcaatcgacgctgc 251

373 GAACCTGCCCAAGCCCATTCACAGTCCGGCTACGTACAGCTGGCCCATGTGACATACGCCAAGCC 432
252 cgtgtggaacacaaagcgaagttgctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 302
433 GTACGGCAAGCATGATCAGGCTCGCGCTGACCGCATCGGATGGGCAACCGGAAACGAGACGA 492
303 ggcattcggaagcactgt 362
493 GCGCGAGGCGCGCTGTGATCTACAAACCGCGCGGCTCGCGGCGCTCGCGGCTTCC 552
363 cgcgcctatgtgcacgc 419
553 GCGCGCGGTACGAAACGAGCGCGGTCTGGGCAACGAGCGCAAGCGCTACGACTTCGT 612
420 ggcgttcgaacccggaagaggttcgc 479
613 CGGCTTGACCG 672
480 cgaattcgacgcgtaacgc 539
673 GTTCGTAAAGCAACCCAAAGCG 732
540 cgtcgaaagcttaccgc 599
733 GCGC---AAGCTGCGCGCGAGTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 789
600 cttggccaatatggtatcgctgc 659
790 CCCGCAATGAC---ACGGCGAACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 846
660 aggtgacatcagaatcaactacactcgc 719
847 GCGGGAAGAAGCTCAACTACCTCGCGCGCTCTCTACGGCACTCTCGCGCGCGCGCGCGCTA 906
720 cctggaagcttcggtatcgtatcgtgc 779
907 GCGGACCTCTTCCGCGGACACG 966
780 cgttagcccaatcgaggaagaagcatcagcaaatgvcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 839
967 CCGGCAAGATCTGTACCAAGCGCAAC-----CTGACGAGACGTGCG 1011
840 ctacgc 889
1012 CTTCGAGGCGCGGTGAAGAGACTGCGGACTGGGTGCGCGGCAACGCGCGCTACCA 1071
900 ggtcaacgcgtaccacgc 959
1072 CTTCGCGGACACCG 1131
960 tcaacgctgtgactgacgc 1019
1132 GAAGAACCGCTGGCGCGCGCTGTGGACCGCGGAGCTGACGACTGATCTCTTTCAGAGCGC 1191
1020 tcagcgctggaagtaacctgaacagtggtctctggtggcgctgcagcgcaagcgcccg 1079
1192 cccgtactagacatcccgcttggtggcgacgacggagatcttcagacaaagtaacgtccggcg 1251
1080 cgaactctgtgtctcttcgcaagactatga---cgccggagatgcagagcgactacag 1136
1252 CGACACCAGCGCTGCTCGACGCGCGGACCGGACCTGTCCGACACCGCGGGCAACGCG 1311
1137 caagcgacaggaagcggttcaacgcggtctcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1196
1312 CTCCGCGGAGAAAGGCAACGCCGCTTACAGCGCGCTGAGTGCAACCGACGCAAGTGCGC 1371
1197 agcgacctgtgtgtgcgcgcgcgaacagatctcagatgcgcgcgcgcgttcttactagcgcgc 1256
1372 CGCAACTGGGCGACCTGGGACCGGGAACACCGCGGCTCCAGCGGACCAACCGCTTCAT 1431
1257 gcaattacccggtatcgcccccgcgatctgtgcgcgtctgtgcgcgtgtgcgcgcgcgcgcgcgc 1316

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 5, 2001, 14:30:29 : Search time 115.27 Seconds
(without alignments)
527.725 Million cell updates/sec

Title: US-09-461-774-4
Perfect score: 2738
Sequence: 1 AAMWRRRLSSALLSFGLL.....AVNHYFLDGLPTSLRCAP 519

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues
Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: SP archaea:*
2: SP bacteria:*
3: SP fungi:*
4: SP human:*
5: SP invertebrate:*
6: SP mammal:*
7: SP mhc:*
8: SP organelle:*
9: SP phage:*
10: SP plant:*
11: SP rodent:*
12: SP virus:*
13: SP vertebrate:*
14: SP unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	780	28.5	539	2	054398
2	653.5	23.9	513	2	054399
3	648.5	23.7	545	2	091000
4	583.5	21.8	529	2	092B05
5	570.5	20.3	537	2	054410
6	547.5	20.0	543	2	069831
7	529	19.3	543	2	09X703
8	379	13.8	499	2	091267
9	298.5	10.9	528	2	P71969
10	261	9.5	526	2	09PB15
11	260.5	9.5	525	2	005685
12	131.5	4.8	433	2	09RD72
13	128	4.7	3122	12	P89459
14	123.5	4.5	1245	2	09RL54
15	123	4.5	1051	2	P70745
16	119	4.3	1079	2	050399
17	117	4.3	374	2	069580
18	117	4.3	3654	2	030766
19	116.5	4.3	585	2	053945

20	116	4.2	9477	2	09L4X3	0914X3 streptomyc
21	115	4.2	92	2	091521	091521 amycolatops
22	115	4.2	447	2	086673	086673 streptomyc
23	115	4.2	532	2	09L105	09L105 streptomyc
24	115	4.2	7525	2	09K1E0	09K1E0 streptomyc
25	114	4.2	1400	2	P96419	P96419 mycobacteri
26	112.5	4.1	274	4	013855	013855 homo sapien
27	112.5	4.1	1603	5	09NE14	09NE14 leishmania
28	112.5	4.1	1815	2	049940	049940 mycobacteri
29	112.5	4.1	7576	2	092G44	092G44 streptomyc
30	112	4.1	643	2	P71750	P71750 mycobacteri
31	111.5	4.1	2129	12	09J6K9	09J6K9 rubella vir
32	111	4.1	2591	2	054959	054959 streptomyc
33	110.5	4.0	2082	2	09S200	09S200 streptomyc
34	110	4.0	429	2	050004	050004 mycobacteri
35	109.5	4.0	282	2	092159	092159 pseudomonas
36	109.5	4.0	524	2	09K3Z7	09K3Z7 streptomyc
37	109.5	4.0	1413	5	09N9N9	09N9N9 leishmania
38	109	4.0	725	2	069970	069970 streptomyc
39	109	4.0	1055	2	053348	053348 mycobacteri
40	108.5	4.0	383	2	09RRH6	09RRH6 deinococcus
41	108.5	4.0	574	2	09RTJ3	09RTJ3 deinococcus
42	108.5	4.0	751	2	092A45	092A45 streptomyc
43	108	3.9	525	2	09RTY1	09RTY1 deinococcus
44	108	3.9	588	2	09XAH5	09XAH5 streptomyc
45	108	3.9	815	10	064481	064481 arabidopsis

ALIGNMENTS

RESULT 1
ID 054398 PRELIMINARY: PRT: 539 AA.
AC 054398;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE PROTEINASE.
GN SLIP OR SCH5.02C.
OS Streptomyces lividans, and Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1916, 1902;
RN [1]
RN [1] SEQUENCE FROM N.A.
RP SPECIES=S. lividans; STRAIN=66;
RC MEDLINE=96042075; PubMed=7592364;
RX Binnie C., Butler M.J., Aphale J.S., Bourgault R., Dizonno M.A., Krygman P., Liao L., Malczyk E., Malek L.T.;
RA "Isolation and characterization of two genes encoding proteases associated with the mycelium of Streptomyces lividans 66.";
RT J. Bacteriol. 177:6033-6040(1995).
RN [2]
RN [2] SEQUENCE FROM N.A.
RP SPECIES=S. coelicolor; STRAIN=A3(2);
RC Seeger K.J., Harris D.;
RX Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RN [3] SEQUENCE FROM N.A.
RP SPECIES=S. coelicolor; STRAIN=A3(2);
RC Bentley S.D., Parkhill J., Barrett B.G., Rajandream M.A.;
RX Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN [4]
RN [4] SEQUENCE FROM N.A.
RP SPECIES=S. coelicolor; STRAIN=A3(2);
RC MEDLINE=97000351; PubMed=8843436;
RX Klenzbach M., Kieser H.M., Denapate D., Eichner A., Cullum J., Kinasli H., Hopwood D.A.;
RA "A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RT Mol. Microbiol. 21:77-96(1996).
DR EMBL, LA2758; AAA93056.1; -.

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RA Redenbach M., Kleser H.M., Denaparte D., Eichner A., Cullum J.,
 RA Kinashi H., Hopwood D.A.;
 RT "A set of ordered cosmids and a detailed genetic and physical map for
 RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.",
 RM Mol. Microbiol. 21:77-96(1996).
 DR EMBL; AL160331; CAG77335.1; -.
 KM Hypothetical protein.
 SQ SEQUENCE 545 AA; 56801 MW; 2E70670C3FF7BEEE CRC64;

Query Match	23.7%	Score 648.5;	DB 2;	Length 545;
Best Local Similarity	34.6%;	Pred. No. 3.3e-39;		
Matches 193;	Conservative 64;	Mismatches 228;	Indels 73;	Gaps 21

[illegible]

RESULT	4		
09ZBJ5			
ID	09ZBJ5	PRELIMINARY;	PRT; 529 AA.
AC	09ZBJ5.		
DT	01-MAY-1999	(TREMBLrel. 10, Created)	
DT	01-MAY-1999	(TREMBLrel. 10, Last sequence update)	
DT	01-JUN-2000	(TREMBLrel. 14, Last annotation update)	
DE	PUTATIVE SECRETED PEPTIDASE.		
GN	SC9C7.15C.		
OS	Streptomyces coelicolor.		
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteriidae;		
OC	Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.		
OX	NCBI_taxid=1902;		
RM	[1]		

RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Seeger K.J., Harris D.;
RL Submitted (Jan-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Bentley S.D., Parkhill J., Barrett B.G., Rajadream M.A.;
RL Submitted (Jan-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kisser H.M., Denapatte D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL: AL035161; CAA2727.1; -.
DR INTERPRO: IPR000073; -.
DR INTERPRO: IPR000379; -.
DR PFMF: PF00561; adenylosase; 1.
SQ SEQUENCE 529 AA: 56400 MW: 6899A076C593C10F CRC64;

Query Match	21.3%	Score	583.5	DB	2	Length	529
Best Local Similarity	32.38%	Pred. NO.	1.6e-34				
Matches 170; Conservative	66	Mismatches	229			Indels	61
						Gaps	19

OY	7	RLLSSALLSFGILLGSLIAAPPLAGATREEPACQTPGA PYVAPOOSNNSCREFIADTSE	66
Dd	2	KORAAVLGCAGAAVIAAGTYAVP--ADASAPRAPHAPAAP--AAGLIIMKCA-----TED	52
OY	67	IIFTARCATVSVDVDQDEGTQAKLAVLEVPAT-GQREGALLVNEGPGASAVDM---VA	122
Dd	53	YPTLQCASILEVFLDHARPGRGRITLALSHVPTACTGYQGPIILVNNGPGSGSLTLAGFA	112
OY	123	AAPPAIADTDILRHNDLVGFDPFRGVGHSTPALRCTDAEFDAVRRDPMADYSFAGTYIVE	182
Dd	113	SSLPA----RVAAQDYDVIGEDPRGVGRSTRPADC-LPHGFDFVRPDSLPD-----TPDVE	162
OY	183	QVYRRLADQCVRMKFSEF---LANIGTSASVARDMMVRALDDOINVLGYSYTKLCTA	239
Dd	163	QAHLARAEFPADAGEBKAYADVLRITYIDYAADADVVALRALGAELKNFTFGYSTGYLLGAV	222
OY	240	YIEREGTHVAMVLDGALDPAVSPLEEESIQMACFOTAFND-----YAADCARSAPCLG	294
Dd	223	YAKLHPDRVRRRLVTVDSVGPDDWVDANLNO---DLAFNDRHRAFLMIAREHDAAVYG	278
OY	295	TDSAQMVNVYYHALVPLVOKP--GKTSIDRGLSYADATGTINALYSP-----QRMKYLTF	347
Dd	279	SPPREVAAWVMYMKRALKKRPAGKV----GASELEDT-----YMGGGYNGWPILA	327
OY	348	SGLLGLQGSADGILLVLADYDGHDADGHSNDODAFENAVRCVYAAPPADPAVVAADO	407
Dd	328	EAFASYAKDEDGPLVEAYENFGALDASG--DNQSVYTAVOCRDPAAMPBMDDEMRDSDW	385
OY	408	RIRQVAFPLSYQQFPCAPADLCALMPVPATSTPPAPAPAGAKVVVVSTTHDPATPYOS	467
Dd	386	DLYEKAPEPVNMANNMNP---CAPWPT-GSQOPADVANGELPAPULLLOQTGDAAFTPEG	441
OY	468	GVDLARQL--GAPLTIFDGTQHTAVFDGNQCVDSAVMHFIJDGTLP	511
Dd	442	GVAARRLLRDSALVVEEGGNGGITLGSNDCCIDEHLAAVLTLDGTVP	487
RESULT	5		
ID	054410	PRELIMINARY;	PRT; 537 AA.
AC	054410:		
DT	01-NOV-1996	(TREMBLrel. 01, Created)	
DT	01-NOV-1996	(TREMBLrel. 01, Last sequence update)	
DT	01-OCT-2000	(TREMBLrel. 15, Last annotation update)	

RESULT	5	
054410		
ID	054410	PRELIMINARY;
AC	054410;	PRT; 537 AA.
DT	01-NOV-1996	(TREMblrel. 01, Created)
DT	01-NOV-1996	(TREMblrel. 01, Last sequence update)
DT	01-OCT-2000	(TREMblrel. 15, Last annotation update)

```

DE TRIPEPTIDYLAMINOPEPTIDASE PRECURSOR.
OS Streptomyces lividans.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
ON NCBI_TaxID=1916;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=96090168; PubMed=7487044;
RA Butler M.J., Binnie C., Dizonno M.A., Krygsman P., Soltes G.A.,
RA Soostmeier G., Walczyk E., Malek L.T.;
RT Cloning and characterization of a gene encoding a secreted
RT tripeptidyl aminopeptidase from Streptomyces lividans 66."
RL Appl. Environ. Microbiol. 61:3145-3150(1995).
DR EMBL: L27466; AAA92338.1; -.
DR MEROPS: S33.002; -.
DR INTERPRO: IPR000073; -.
DR INTERPRO: IPR000734; -.
DR PFAM: PF00561; abhydrolase_1.
DR PROSITE: PS00120; LIPASE_SER; UNKNOWN_1.
DR Signal: Aminopeptidase.
SIGNL 1 39 POTENTIAL.
CHAIN 40 537 TRIPEPTIDYLAMINOPEPTIDASE.
SQ SEQUENCE 537 AA; 58273 MM; 850703374BC4DEC9 CRC64;

Query Match 20.8%; Score 570.5; DB 2; Length 537;
Best Local Similarity 31.1%; Pred. No. 1.5e-33;
Matches 173; Conservative 82; Mismatches 233; Indels 69; Gaps 21;

QY 5 RRRPLSSALLSFGLLGLLAAAPPLAGATEBPAGAGTP-----GAPVVA 49
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 2 RKSSIRRRATAFG--TAGALVTATLTAAGVSAAPASAPADGHHGHSMDREARAAIAA 59
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 50 PQOS-----WNSCREPLADTSEIRTARCATVSPVDYDQCGTQAKLAVRVATG---Q 101
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 60 ARAARAGIDMEDC--AADWNLPKPIQCGYTVPMDYAKPYGQIRLAVRIGNTGRSE 116
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 102 RFALLVNPFGPGASAVDWAAMAAPAIAD-TDILRHFDLVGFPPRGHSTPALRCRTDA 160
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 117 RQALILYNPGPGSGSLRFARVARTNKSAAVMAANTAKADYFGFPRGCGHAP--ISCVDQ 175
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 161 EF-DAYRRDPMADYSPAGVTHVEQVYRQLAODCVDRMGESFLANIGTASVARDMDVROA 219
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 176 EFVKAKRKPADVPG--SEADKRAQRKLAREVAGCFERSG--EMLPHMTTPNTARDLDVIRA 233
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 220 LGDDQINYLGSYGTGKTGYLERFQTHVRAMLDGALDAVAVPI--EESISQAGQTA 277
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 234 LGKKKLNYLGVSYGTYLGAVYGTLPFDHVRMVDVSVNRSRDKIYQANLDDVAEGR 293
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 278 FNDYAA-DCARSPACPLGTDSAQMVNRYHALVPLVOKP--GKTSDPKGL-----SYAD 328
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 294 WKWQMDVMAANDAAHYHGDTRAEVQDQWLKLRAAAKKPLGIVVGPARELISFQSAPIYD 353
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 329 ATTGTINALYSPOKWKYLTSGLLGLQSGDAGDLVLADY--DGRDADGHS--NDQDA 384
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 354 SAAPPAETFS---KYV-----AGDTQALVDAAPDLSPDASNAASENGNAV 397
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 385 FNNVRCVYVPTPADPAWVAADQRIQVAPFLSYGQFTGSAPRDLCALWVVPATSTPHRA 444
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 398 YTAVECTDAWPMANMRTWDRNTRLRHDFPFMTWANAAMNLP--CATWVVKQOTPLNVK 454
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 445 APAGACKVVVVSTTHDPATPYGSGVDLARQL-GAPLIT-FDGTQHTAVFDGNCQVDSAWA 502
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 455 TGGGLGSPVLIVGSEKDAATPYEGAVELHOFKRSRLITTEADAGSHGVTGLVNPCLINDRYD 514
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 503 HYFLDGLTLPPTSLRCAP 519
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 515 FYLLTGRTARDVTCAP 531
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

AC 069831;
DT 01-AUG-1998 (Tremblrel. 07, Created)
DT 01-AUG-1998 (Tremblrel. 07, Last sequence update)
DT 01-MAY-2000 (Tremblrel. 13, Last annotation update)
DE PUTATIVE SECRETED PROTEASE.
GN SCIB5.03C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
ON NCBI_TaxID=1902;
RN (1)
RP SEQUENCE FROM N.A.
RX STRAIN-A3(2);
RA Harris D., Taylor K.;
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
RN (2)
RP SEQUENCE FROM N.A.
RX STRAIN-A3(2);
RA Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
RN (3)
RP SEQUENCE FROM N.A.
RX STRAIN-A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denaplatte D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome."
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL: AL023517; CAI18977.1; -.
DR INTERPRO: IPR000073; -.
DR PFAM: PF00561; abhydrolase_1.
DR Protease.
SQ SEQUENCE 543 AA; 58514 MM; 2D3E35E68E61A8B8 CRC64;

Query Match 20.0%; Score 547.5; DB 2; Length 543;
Best Local Similarity 32.0%; Pred. No. 7e-32;
Matches 178; Conservative 71; Mismatches 221; Indels 87; Gaps 24;

QY 10 SSALLS-FGLLGLLAAAPPLAGATEBPAGAGTP-----WNSC--REF 61
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 13 AALYVSAAGSLTATLTAAVRAAPRAGPAPALCTAAVAAARRAAAGVDFGCGAEEL 72
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 62 ADTSEIRTARCATVSPVDYDQCGTQAKLAVRVAT-----GQRFGLVNA 109
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 73 PD-----GTRCGTIVTPLDANPDGKQVRLTVSRVATHRPDGGGREVPGQ--GALLIN 125
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 110 PGPGASAVDM-VAAAPAIADTDILRHFDLVGFDPGVSHTPALRCRTDAEF----- 162
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 126 PGPGASGLYFPLVGVPEW--KRIMAAVDLGVARGVAPSAF--LSCQDPGDFKGPFR 182
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 163 -----DAYRRDPMADYSPAGVTHVEQVYRQLAODCVDRMGESFLANIGTASVARDMDVROA 216
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 183 APTRPSPDAYKKERAAE-----AEAY---ARCDAR-KGGEALRHYSLNAAARLEVV 229
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 217 RQALGDDQINYLGSYGTGKTGYLERFQTHVRAMLDGALDAVAVPI--EESISQAGF 274
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 230 RAALGERRLTFKASYGTYLGALYLAELFPSHVRRMVFDAVAVNPDPQVWYRNNDQSEAF 289
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 275 QTAFFNDYAADCARSPAC-PLGTDSAQMVNRYHALVPLVOKP--GKTSDPKGLSYADATG 332
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 290 EDWMDPFKEVVAEHDAVYGLGTRARQVODSYERASRLAAEPAGTGVGAQGTGALITAG 349
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 333 TINALYSPOKWKYLTSGLLGLQSGDAGDLVLADYDGRDADG--HYSNDQDAFNAVRC 390
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 350 YYDD-YWPHRAEALSAVL-----KGDPOPLVDAGRTAGAVRSENSNAVYTAVEC 400
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 391 VYAPTPADPAWVAADQRIQVAPFLSYGQFTGSAPRDLCALM-----PVPAITSTPHRA 444
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 401 NDAWPMPADEVMDRDNTRLARRAFETWNVWNTLP--CAVWQGPORPRDLRTAPGEL 457
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 445 APAGACKVVVVSTTHDPATPYGSGVDLARQL-GAPLIT-FDGTQHTAVFDGNCQVDSAWA 502
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

Db 458 PP-----TLVLAERDAAPYEGALELRRLRAGSVLTERDAGTGIAGGPACVNGHLE 512
 QY 503 HFEIDGTLPRTSLRCAP 519
 Db 513 AYLDGRLPARDTSCAP 529

RESULT 7

09X7U3 PRELIMINARY: PRT: 543 AA.
 AC 09X7U3:
 DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
 DE PUTATIVE PROTEASE.
 GN SC5H1.36.
 OS Streptomyces coelicolor.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
 NCBI_TaxID:1902;
 [1]

RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RA Oliver K., Harris D.;
 RT "A set of ordered cosmid and a detailed genetic and physical map for
 the 8 Mb Streptomyces coelicolor A3(2) chromosome."
 RL Submitted (MAY-1999) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RA James K.D., Parkhill J., Barrell B.G., Rajandream M.A.;
 RL Submitted (MAY-1999) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RA Redenbach M., Kieser H.M., Denapalte D., Eichner A., Cullum J.,
 RA Kinashi H., Hopwood D.A.;
 RT "A set of ordered cosmids and a detailed genetic and physical map for
 the 8 Mb Streptomyces coelicolor A3(2) chromosome."
 RL M.O. Microbiol. 21:77-96(1996).
 DR EMBL: AL049863; CAB42961.1; -
 DR INTERPRO: IPR000073; -
 DR INTERPRO: IPR000379; -
 DR PFAM: PF00561; abhydrolase; 1.
 KW Protease.
 SO SEQUENCE 543 AA; 57696 MW; DA6549032B9F4B6 CRC64;

Query Match 19.3%; Score 529; DB 2; Length 543;
 Best Local Similarity 29.2%; Pred. No. 1.5e-30;
 Matches 166; Conservative 76; Mismatches 227; Indels 100; Gaps 19;

QY 6 RRRPSSALLSFGLLGLLAAPLAGA-TEBPAGOTPGAVVAPQOSWNSCREFIADT 64
 Db 12 RRCATVAGLALGALGAPAAASGATTTATRTGSGL-----SFFYORVEMTACE-----GT 64
 QY 65 SEIRTRCATVSPVVDYDQPGTQAKLAVIRVPATGQRFGLLVNPGFGASAVDMAAM 124
 Db 65 EMPKDLQCGKVVPLDYSRPRGCTLDVALARYRATGDSGSLVLFNGGGSGINELALG 124
 QY 125 APAIADTILRHFDLVGDPDPRGVHSTP-----ALRCTDAEFAYRRDPADYSPAVY 178
 Db 125 GKFF--MHITNGYDVVTDPDPRGVSRSSPVSCGPATLKIMEATDGDGMDTDPG----- 175
 QY 179 THVEGVYRQLADQVDRMGFSFLANIGTASVARMDMYRQALGDOQINTLGSYGKILGT 238
 Db 176 ---VLERLDAAECAKAKYSG-PVLPHITGYDARMDYRRALGDRLNTLNGSYGTRICA 232
 QY 239 AYLRFEGTHVRAMVLGDAIDPAVSPV-EESSISQMAFGOTAFNDYAADCARSPACPLGIDS 297
 Db 233 VYAAFRFPKVGKRVLDG-VDTLTERPLAEQGLAGARGCGQALENFLDMCVEDVACPFQQA 291

QY 298 AQWNRHYALVDLYOKPEKTSDPRLGSLYADATGTI-----NALYSPQRRKYL----- 346
 Db 292 RDARQOVERLVA SL-----DSDPYPSAFGEPTTQDMYGAIGQALYSELMPSLEKALA 345
 QY 347 -----TSGLLGIQSGSDAGDLVLADYDGRDADGHYSNDQ-----AFNAVRCV 391
 Db 346 OLLEDGDFRGLGECFSSGCVTPVRAAVNAPVGRDTRAGLTDEEDVPMNLPAALMAINCA 405
 QY 392 YAPTADPAAWVAADQRI R-----QVAP-----ELSYGQFTGS-----APRDLCAL 432
 Db 406 DDPDRPTAAQVYASIDRLNARAYEDVSPVGRYRLQVLMCYGRPRGTPTYIRDDVDL--- 462
 QY 433 WVPAFTSPRHPRAAGAGGVVVVSTTHPPAPRYOSGVDLAROLGAPLITFD--GTQHTAV 490
 Db 463 -----DTABMLVGTGRDPAIPYRMTTETADRLGPSAVVLNRRGCHTG- 506
 QY 491 PDGNOCVDSAVMHYFDLTLPPTSLRCAP 519
 Db 507 YASSKCVHRKXVDFFLYGSLPPDSSCGP 535

RESULT 8

09L2G7 PRELIMINARY: PRT: 499 AA.
 AC 09L2G7:
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
 DE PUTATIVE SECRETED PROTEIN.
 GN SCG121.19.
 OS Streptomyces coelicolor.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
 NCBI_TaxID:1902;
 [1]

RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RA Oliver K., Harris D.;
 RL Submitted (JAN-2000) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
 RL Submitted (JAN-2000) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RA Redenbach M., Kieser H.M., Denapalte D., Eichner A., Cullum J.,
 RA Kinashi H., Hopwood D.A.;
 RT "A set of ordered cosmids and a detailed genetic and physical map for
 the 8 Mb Streptomyces coelicolor A3(2) chromosome."
 RL M.O. Microbiol. 21:77-96(1996).
 DR EMBL: AL137166; CAB69737.1; -
 SO SEQUENCE 499 AA; 53138 MW; 4553D0B88093E753 CRC64;

Query Match 13.8%; Score 379; DB 2; Length 499;
 Best Local Similarity 27.2%; Pred. No. 1.1e-19;
 Matches 146; Conservative 78; Mismatches 245; Indels 68; Gaps 22;

QY 5 RRRPSSALLSFGLLGLLAPLAGATEBPAGOTPGAVVAPQOSWNSCREFIADT 64
 Db 6 RSRSTRNAI-VGVSTALGVTAAPVTA GA-----AGGSTTDA-----LTWSACBGTGLDP 54
 QY 65 SEIRTRCATVSPVVDYDQPGTQAKLAVIRVPA--TGORFGLLVNPGFGASAVDMAA 122
 Db 55 RQ-----ECATLDVPMADYDPPGRIEMASRTPAKPAKRRALLLNGCGSSLNPS 110
 QY 123 A---MAPAIADTILRHFDLVGDPDPRGVHSTPALRCRT--DAEFDAYRRDPADYSP 175
 Db 111 GKGQRLPQEVFRDT-----HDLIGFAPRGLAPSTPA-DCGLEYGATLATSRLRPWAPPDGSV 164

RESULT	9			
P71969		PRELIMINARY;	PRT;	528 AA.
ID	P71969			
AC	P71969;			
DT	01-JAN-1998 (TrEMBLrel. 05, Created)			
DT	01-JAN-1998 (TrEMBLrel. 05, Last sequence update)			
DT	01-JUN-2000 (TrEMBLrel. 14, Last annotation update)			
DE	HYPOTHETICAL 54.0 KDA PROTEIN.			
GN	RV2672 OR MTCY441.41.			
OS	Mycobacterium tuberculosis.			
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;			
OC	Actinomycetales; Corynebacterineae; Mycobacterium.			
OX	NCBI_TaxID=1773;			
OX	[1]			
RN				
RP	SEQUENCE FROM N.A.			
RC	STRAIN=H37RV;			
RA	Skellon J., Churcher C.M., Barrell B.G., Rajandream M.A.;			
RA	Submitted (SEP-1996) to the EMBL/GenBank/DBD databases.			
DR	EMBL; 280225, CAB03326.1; -			
DR	TUBERCULIST; RV2672; -			
DR	INTERPRO: IPR000073; -			
DR	PFAM; PF00561; abhydrolase; 1.			
DR	Hypothetical protein.			
SEQUENCE	528 AA; 54015 MW; 64FF0A1A9701668 CRC64;			

```

QY 7 RPISALLSEFLLIGGLLLA-----PPLGAEIEEPACGPGAPV 48
    : : : : :
Db 9 RPMSTAMVLVALTGSATVLAACVPACFADPREATVSGACPGQAATTTPPAGP---PPLA 65

QY 49 APQO--GWSNCREFIADTSERTA-----RCATVSVYVDKDPGCGQAKLAVIRVPA--T 99
    : : : : :
Db 66 APKMDLSMHDSITSNVYVSNAGTIPAPGYKLECASTDITDLDPELVGGSTAVSIGVYRANSNOT 125

QY 100 GQRFALLVNPGGGASAVDMVAAMADIADTDILRHFDLVEFGDPBGVGHSTPALRCRTD 159
    : : : : :
Db 126 PSDAGRLVFTTSGDLPSPSTOLPWLH--AGIDVLRSHPIVAADRMRGMSGSP-IDCR-- 180

QY 160 AEFPAUYRRDPADVS--PAG--VTNHEOVYRQDAOCQVVRM--GFSFLANIGTASVARD 212
    : : : : :
Db 181 ---DHFRDEKRDQAQQAQDDPRANISDTSNATTTCTTALAPGESAUYNTNHAAS---D 234

QY 213 MDWVRQAAGDDQINLYGYSYGTKLGTAYLERFEGTHVRAMVLDGALDPVSPVIEESISQMA 272
    : : : : :

```

RESULT	10			
09PB15				
ID	09PB15	PRELIMINARY;	PRT;	526 AA.
AC	09PB15;			
DT	01-OCT-2000 (Tremblrel. 15, Created)			
DT	01-OCT-2000 (Tremblrel. 15, Last sequence update)			
DT	01-OCT-2000 (Tremblrel. 15, Last annotation update)			
DE	PROTEINASE.			
GN	XP2330.			
OS	Xylella fastidiosa.			
OC	Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;			
OC	Xylella.			
OX	NCBI_TaxID=2371;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	SPRAIN-9A5C:			
RX	MEDLINE=20365717; PubMed=10910347;			
RA	Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,			
RA	Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,			
RA	Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Britons M.R.S.,			
RA	Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrier H.,			
RA	Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,			
RA	Costinho L.P., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,			
RA	Falcinelli A.P., Ferreira A.J.S., Ferreira V.C.A., Ferio J.A.,			
RA	Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,			
RA	Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,			
RA	Ho P.L., Honeisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,			
RA	Krieger J.E., Kuranae E.E., Laigret F., Lambais M.R., Leite L.C.C.,			
RA	Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,			
RA	Machado M.A., Madella A.M.B.N., Madella H.M.F., Marino C.L.,			
RA	Marques M.V., Martins E.A.L., Martins E.M.F., Matsushima A.Y.,			
RA	Merck C.F.M., Miraca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,			
RA	Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,			
RA	Miani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,			
RA	de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,			
RA	Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,			
RA	Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,			
RA	de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,			
RA	da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,			
RA	da Silveira J.F., Silvestri M.L.Z., Silveira W.J., de Souza A.A.,			
RA	de Souza A.P., Terenzi M.F., Truffi D., Tsai S.W., Tsubako M.H.,			
RA	Valada H., Van Sluys M.A., Verjovsky-Almeida S., Vettore A.L.,			
RA	Zago M.A., Zatz M., Meidanis J., Setubal J.C.;			
RT	"The genome sequence of the plant pathogen Xylella fastidiosa.";			
RL	Nature 406:151-157(2000).			
DR	EMBL: AEC004043; AAF85129.1; -			
DR	INTERPRO: IPR000073; -			
DR	INTERPRO: IPR000379; -			
DR	INTERPRO: IPR000734; -			
DR	INTERPRO: IPR002410; -			

UA NCBI_taxid=1902;
 RN [1]
 RP
 RC
 SEQUENCE FROM N.A.
 STRAIN=A3(2);
 SA Saunders D.C., Harris D.;

RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [12]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Thomson N.R., Parkhill J., Barrell B.G., Rajandream M.A.,
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [13]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Redenbach M., Kieser H.M., Denapate D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL: AL136502; CAB6205.1; -;
DR INTERPRO: IPR000073; -;
DR INTERPRO: IPR000379; -;
DR INTERPRO: IPR002410; -;
DR PFAM: PF00561; abhydrolase; 1.
DR PRINTS: PR00793; PROAMNOPTASE.
DR Aminopectidase.
SQ SEQUENCE 433 AA; 47580 MW; EF9DA0344C10EF55 CRC64;

Query Match 4.8%; Score 131.5; DB 2; Length 433;
Best Local Similarity 23.6%; Pred. No. 0.084;
Matches 105; Conservative 42; Mismatches 165; Indels 133; Gaps 22;

QY 76 SVYVDVDPGCTGQAKLAVINPAT---GGRFGLLVNPGPGASAVVMAAPAIADTD 132
18 TVPLDHAADPGETELAYAREAVASDKADQDLPMLVYLQGGPGGKADRFVG--RPAMFGR- 74
QY 133 ILRHFDLVGDPGCVGSHSTALRCRTDAEFDAYRRDP--MADYSPAGVTHEQYVROLAQ 150
75 ALMEYRILLDDQGTGASTASR-----QTLPLRGGADEADY----- 112
QY 191 DCVDRMGESFLANIGTASVARDMDNR-QALGDDQINLYGSGYGTGLGTYLERFGTHVR 249
113 -----LAHFRADAIYRDECAIRPQYTGAPMTVLQSGFGCTVAYVLSLAPEGLS 162
QY 250 AMVLDAIDRAVSPIESISOMAGFOFAFNDYADCAKRSRACPLGTSAAQVNRVHALVD 309
163 AALITGL-PSLDAHADVDYR-AAVPRIERKVVANVARYP-----QDVERARRIAD 211
QY 310 PLVQKPKTSD---PRGLSYADATTGTINALYSPQRKYLTLGSLGQSGSDAGD--LLV 364
212 HLL-----TRDVVLPNKYRL-----TVEAFQS-----LGIMLGSGESHLHF 249
QY 365 LADYDGRDAGHYSDQDAFN-----AVRCVYA-----PRPADPAAMVAADQR 408
250 LLDADAFVFRANGHELIS--DAFOEOAGCLSYAGRPYALVHEALYGGDARPTDMSA--ER 305
QY 409 IROYAPFLSYGQ-----FTG-----SAPROLCALM-P 434
306 VRAEFPPDAKAKLAGDEPLFTGESIHPMWFDCCPALRPLRETAELLARTWTPLYDP 365
QY 435 VPATSTPHPAAPAGACKVVVVSTTH 459
Db 366 ARLAANEVPAALAAVYHDDMYVDTAH 390

RESULT 13
P89459 PRELIMINARY; PRT: 3122 AA.
AC P89459;
DT 01-MAY-1997 (Tremblrel. 03, Created)
DT 01-MAY-1997 (Tremblrel. 03, Last sequence update)
DT 01-JUN-2000 (Tremblrel. 14, last annotation update)
DE VERY LARGE TECUMENT PROTEIN.
GN UL36.
OS Herpes simplex virus (type 2).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Simplexvirus.

OX NCBI_TaxID=10310;
RN [11]
RP SEQUENCE FROM N.A.
RC STRAIN-HG52;
RX MEDLINE=87111457; PubMed=3027242;
RA McGeoch D.J., Moss H.W., McNab D., Frame M.C.;
RT "DNA sequence and genetic content of the HindIII 1 region in the short
RT unique component of the herpes simplex virus type 2 genome:
RT identification of the gene encoding glycoprotein G, and evolutionary
RT comparisons.";
RL J. Gen. Virol. 68:19-38(1987).
RN [12]
RP SEQUENCE FROM N.A.
RC STRAIN-HG52;
RX MEDLINE=90278430; PubMed=2161906;
RA Everett R., Fenwick M.;
RT "Comparative DNA sequence analysis of the host shutoff genes of
RT different strains of herpes simplex virus: type 2 strain HG52 encodes
RT a truncated UL41 product.";
RL J. Gen. Virol. 71:1387-1390(1990).
RN [13]
RP SEQUENCE FROM N.A.
RC STRAIN-HG52;
RX MEDLINE=92113549; PubMed=1662697;
RA McGeoch D.J., Cunningham C., McIntyre G., Dolan A.;
RT "Comparative sequence analysis of the long repeat regions and
RT adjoining parts of the long unique regions in the genomes of herpes
RT simplex viruses types 1 and 2.";
RL J. Gen. Virol. 72:3057-3075(1991).
RN [14]
RP SEQUENCE FROM N.A.
RC STRAIN-HG52;
RX MEDLINE=92356101; PubMed=1322965;
RA Barnett B.C., Dolan A., Telford E.A.R., Davison A.J., McGeoch D.J.;
RT "A novel herpes simplex virus gene (UL49A) encodes a putative membrane
RT protein with counterparts in other herpesviruses.";
RL J. Gen. Virol. 73:2167-2171(1992).
RN [15]
RP SEQUENCE FROM N.A.
RC STRAIN-HG52;
RA Dolan A.;
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: 286099; CAB06722.1; -;
SQ SEQUENCE 3122 AA; 330047 MW; 6EBF94B51BFE8C0B CRC64;

Query Match 4.7%; Score 128; DB 12; Length 3122;
Best Local Similarity 22.2%; Pred. No. 2.2; Mismatches 166; Indels 182; Gaps 25;
Matches 111; Conservative 42;

QY 1 AAMWRRRPLSALLSPGLLGLLLAAPLAGATEEPGAGQTPGAPVAPQGSMSNCREF 60
2300 AATTHRTF-AAKYTAADLVLAAYVL-----GAPVV-----ALRNT 2334
QY 61 IADTSEIRARCATVSVYVDYDQPGGTQAKLAVIRPATGREGALVNPFGGASAVDM 120
2335 TARSRESELELCLTL-----FDS-----RPGGDALRDY 2364
QY 121 VAAMAPAIADTDI-----LRHFDLVGFPDRGVGSHSTPALRCRTDAEFAYRRDPADYS 174
2365 VS-----SDIETWAVGLLHTDLPNFINACLAOLPRLISA-----LIARPLADGP 2409
QY 175 PA-----GVTHVEOYV-----ROLAODCVDRMGESFLANIGTASVARDMD--M 215
2410 PCLVLDISMTFPAVILMEAPPGPDVRRVGSSEATEELPFVATAGDVLAASADADPF 2469
QY 216 VROLGDDQDQINLYGSGYGTGLGTYLERFGTHVRAMVLDCAIDPAVSPIESISOMAGFO 275
2470 ARAIL-----GRPDASILTGLPFGHPVYQORPLADEAG-- 2503
QY 276 TAENDVAADCAKSPACPLGDSQAWNRVHALVDPLQKKGKTSDRGLS---YADATTG 332
2504 ---PSAPTAARDPRLAGDGGSGPE-----DP-AAAPARQADPGVLAFTLLTDTTG 2552

Db 194 PVASALVSSSGFSATSTGYVGTSDGTDLADKLDGAYPEAGPGNIQTGRIPLTGAKT 253
QY 53 SWNSCREFIADTSEIFTRCARTVS-----VPVDY-----DQPG-----GTOA 89
Db 254 EFSIALGFGMAEPAVNTANTSVSRGFPKYSKDYTKEMKYLQSLDPKAKALDGELERTQY 313
QY 90 KLAIVIRVPATGOR-----FGALLVNPGGPGASAV-----DMVAMAPATA-D 130
Db 314 DVSIMTYKHAHEDTTPGAWILASLTTPMGQVAMNEQHREGYHAWVARDMYQSVTALLAAGD 373
QY 131 TDILR-----HFDLVGFDPRGVGH-----STPALRC-----RTDAEFD 163
Db 374 TDAARRGVEWLFKYQQQPDGHFPQTSKVDGTGQNGIQLDETAEPILLANQIGRTDADFY 433
QY 164 AYRRDPWADYSPA-----GVTHVEQYVROLAQ-----DCVDRMGFSFLANIG 205
Db 434 AKELKPRADILVAGPFTPOERMEETGYSTILASQIALAALAAAGDIAEKNG-----DVG 488
QY 206 TASVARDMDMYROLGDD-QINYLGYSGYTK-----IGFAYL-----ER 243
Db 489 SAAIYR-----ATADEMQNTEKWMFTNGPYGDNYYLRTISGSGNPNDGAARDWNG 541
QY 244 FCTHVRAMVLDGAI-----DPANSPTEBESIQMAGFOTAFENDYAADCARSPACPLG 294
Db 542 AGVHPENAVLDGGFLEFVLGVKRPADAHVADSIAE-----TDASISQETPGG 589
QY 295 TDSAQWVNRHYALVDPLOKRPKTSDEPRGLSYADATGTINALYSPQRMKYLTSGLIGLQ 354
Db 590 ---RMW-HRY-----IYDGYGEKAD--GSPMDGTGVGRMLPLLSERGEY-----ALA 631
QY 355 RGSDDAGDLVLAD-----DYDGRDADGHYSNDQAFNAVRCVVAETPADPA 400
Db 632 NGODALPYLATMHSANOGFMIPQVWVDDODEPTSYGHELGSTG-----SASPL 680
QY 401 AMWAADQRIQVAPFLSYGQFTGS---APRDLGALW-----PVPATSTPHRAPAPAG 450
Db 681 SWMAA-QYVR-----LAAGITKNSPEVETPKNVQRYASGAVGRKPAITINSPOULTTADSA 734
QY 451 KVVVVSTT-----HDPATPYQSG-----VDLAROLGAPLITFDGTOHTAVF 491
Db 735 NTIVYSGTTAAKAYVSVNGNIFEFAPLTPAGDGTSTFSVDLA-----LPATKNTVTV 765
QY 492 -----DGNOCVDS-AVMHY 504
Db 786 AAVGSDGTAESRTVLHY 804

Arch completed: January 5, 2001, 14:30:37
Time: 329 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 5, 2001, 14:31:05 ; Search time 35.06 Seconds
(without alignments)
478.056 Million cell updates/sec

Title: US-09-461-774-4
Perfect score: 2738
Sequence: 1 AAMWRRRLSSALLSFGILL.....AVMHYFDGTLPTSLRCAP 519

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 88757 seqs, 32294092 residues
tal number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2717	99.2	520	1 YM23_MYCTU	Q10508 mycobacteri
2	1286.5	47.0	520	1 YM24_MYCTU	Q10509 mycobacteri
3	493	18.0	499	1 YUAR_ECOLI	P34211 escherichia
4	128.5	4.7	424	1 PIP_AERSO	P46547 aeromonas s
5	113.5	4.1	459	1 COBG_PSEDE	P21637 pseudomonas
6	105.5	3.9	707	1 PSF_HUMAN	P23246 homo sapien
7	104	3.8	402	1 KAS2_STRCM	P41176 streptomyce
8	103.5	3.8	1461	1 IE18_PPRIF	P11675 pseudorabie
9	102	3.7	2110	1 MCAS_MYCBO	Q02251 mycobacteri
10	101	3.7	359	1 OMPA_SERMA	P04845 serratia ma
11	101	3.7	498	1 AMYA_ASPAW	Q02905 aspergillus
12	101	3.7	499	1 AMYB_ASPAW	Q02906 aspergillus
13	99.5	3.6	446	1 LYS9_YEAST	P38999 saccharomyc
14	99.5	3.6	466	1 STL_MYCTU	P71698 mycobacteri
15	99	3.6	1632	1 CRPI_MYCTU	Q10900 mycobacteri
16	98.5	3.6	614	1 ODP2_HUMAN	P10515 h dihydroli
17	98.5	3.6	621	1 TLVB_MYCAV	Q59498 mycobacteri
18	98.5	3.6	806	1 NDOG_MYCTU	P95175 mycobacteri
19	98	3.6	314	1 BLA2_STRCI	P14560 streptomyce
20	97	3.5	1162	1 TCNA_TRYCR	P23253 trypanosoma
21	96	3.5	674	1 BAT_HALSA	P13260 halobacteri
22	95.5	3.5	941	1 CGSP_MYCTU	Q50601 mycobacteri
23	95.5	3.5	1741	1 CO4_HUMAN	P01028 homo sapien
24	95.5	3.5	2208	1 POLN_MANCV	Q69014 manchester
25	94.5	3.5	486	1 VG14_BPMI5	Q05220 mycobacteri
26	94	3.4	362	1 RLPA_ECOLI	P10100 escherichia
27	93.5	3.4	483	1 DMPD_PSESP	P19076 pseudomonas
28	93.5	3.4	473	1 LEU2_MYCTU	Q53337 mycobacteri
29	93.5	3.4	619	1 CHIT_STRLI	P36609 streptomyce
30	93	3.4	499	1 AMYA_ASPOR	P10529 aspergillus
31	93	3.4	541	1 Y4MM_RHISN	P55572 rhizobium s
32	93	3.4	547	1 CP78_MAIZE	P48420 zea mays (m
33	93	3.4	644	1 ARSA_HALSA	O52027 halobacteri

34	93	3.4	1848	1 CBPA_CLOCL	P38058 clostridium
35	92.5	3.4	463	1 Y102_MYCTU	O53951 mycobacteri
36	92.5	3.4	703	1 U117_HSV11	P10201 herpes simp
37	92.5	3.4	1912	1 CHD4_HUMAN	Q14839 homo sapien
38	92	3.4	392	1 NYDB_FLASK	P07061 flavobacter
39	92	3.4	456	1 REP_STRLI	P22406 streptomyce
40	92	3.4	1046	1 MEXB_PSEAE	P52002 pseudomonas
41	92	3.4	1235	1 DPOL_HSV11	P04293 herpes simp
42	92	3.4	1235	1 DPOL_HSV1S	P09854 herpes simp
43	91.5	3.3	437	1 YTH3_RHOER	P46372 rhodococcus
44	91.5	3.3	474	1 NOEK_RHISN	P55356 rhizobium s
45	91.5	3.3	495	1 UXNA_ECOLI	P42604 escherichia

ALIGNMENTS

RESULT 1	ID	YM23_MYCTU	STANDARD:	PRT:	520 AA.
AC	Q10508:				
DT	01-OCT-1996 (Rel. 34, Created)				
DT	01-OCT-1996 (Rel. 34, Last sequence update)				
DT	30-MAY-2000 (Rel. 39, Last annotation update)				
DE	HYPOTHETICAL. 55.1 KDA PROTEIN RV2223C PRECURSOR.				
GN	RV2223C OR MRCY427.04C.				
OS	Mycobacterium tuberculosis.				
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;				
OC	Actinomycetales; Corynebacteriineae; Mycobacteriaceae; Mycobacterium.				
RN	[1]				
RC	SEQUENCE FROM N.A.				
RP	STRAIN-H37RV;				
RX	MEDLINE:98295987; PubMed-9634230;				
RA	Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,				
RA	Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,				
RA	Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,				
RA	Davies R., Devlin K., Fellwell T., Gentles S., Hamlin N., Holroyd S.,				
RA	Hornsby T., Jagsels K., Krogh A., McLean J., Moule S., Murphy L.,				
RA	Rutter S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,				
RA	Taylor K., Whitehead S., Squires S., Squires R., Sulston J.E.,				
RT	"Deciphering the biology of Mycobacterium tuberculosis from the				
RL	complete genome sequence."				
CC	Nature 393:537-544(1998).				
CC	-I- SIMILARITY: TO E.COLI YZZE AND M.TUBERCULOSIS RV2224C.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@sib-sib.ch).				
CC	-----				
DR	EMBL: 270692; CA94646.1; -				
DR	MEROPS: S33.006;				
DR	TUBERCULIST: RV2223C; -				
DR	INTERPRO: IPR000073; -				
DR	PFAM: PF00561; abhydrolase; 1.				
KW	Hypothetical protein; Transmembrane; Signal.				
FT	SIGNAL 1 34				
FT	CHAIN 35 520				
FT	TRANSMEM 104 124				
FT	SEQUENCE 520 AA: 55077 MW; 7D05F153A4C6B91 CRC64;				

Query Match Best Local Similarity 99.4%; Pred. No. 9.9e-185;

Matches 516; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAMWRRRLSSALLSFGILLGGLLAAPPLAGATEEPGAGQTPGAPVVAPOGSWNSCRFF 60
|||||
DB 2 AAMWRRRLSSALLSFGILLGGLPLAAPPPLAGATEEPGAGQTPGAPVVAPOGSWNSCRFF 61


```

RT "Complete nucleotide sequence of the F plasmid: its implications for
RT organization and diversification of plasmid genomes."
RT Submitted (Apr-2000) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE OF 1-226 FROM N.A.
RC STRAIN-K12 / ATCC 12435.
RC MEDLINE=94117370. PubMed=828530;
RA Kaufmann A., Slierhof Y.-D., Hennig U.,
RT New outer membrane-associated protease of Escherichia coli K-12."
RL J. Bacteriol. 176:359-367(1994).
CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S33.
CC -I- SIMILARITY: TO M.TUBERCULOSIS RV2223C AND RV2224C.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X74278; CAA52339.1; -.
DR EMBL: AP001918; BAA97900.1; -.
DR PIR: S37474; S37474.
DR ECOGENE: EG40043; Y0AR.
DR INTERPRO: IPR000073; -.
DR INTERPRO: IPR002410; -.
DR PFM: PF00561; abhydrolase.1.
DR PRINTS: PR00793; PROAMNPTASE.
DR Hypothetical protein, Hydrolase, Signal; Plasmid.
KM SIGNAL 1 26
FT CHAIN 27 499
FT CONFLICT 418 499
FT FT
FT FT
FT FT
SO SEQUENCE 499 AA: 55682 MW: 61258807134A26A CRC64;

Query Match 18.0%; Score 493; DB 1; Length 499;
Best Local Similarity 26.7%; Pred. 1e-27;
Matches 131; Conservative 96; Mismatches 199; Indels 66; Gaps 16;

OY 47 VVAPQSNWSC-----REFIADTSEIFRARCATVSPVYDQPGGTQA-----KLAVI 94
DB 25 VSARQIEWQSCWTSPIYSDWFGKSSPELLCGYLSPVKYTFDTGQVSDENIPVLAMT 84
OY 95 RVPATGQRFAGLLVNPGGPQASVDMVAAMPAT---ADTDILRHFDLVGDPQGVGHS 150
DB 85 KLPASKRKRGVYIIISGGGLPGIN-----FYINFDMPVTNLRSSWPIIGDPQGVQS 138
OY 151 TPALRCRFDPAEDFAYRBPMDADSPAGVTHVQVYROLAODCDVDMGSPFLANTISAVA 210
DB 139 FPAINCQS-----NGERLVNVS-----KQLLKQKINACINHTGAEVITHISHEAV 186
OY 211 RDMDMVROLGDDQINVLGYSTYKLGTAUYLERFGTHVAMVLDGADIPAVSIEESIS- 269
DB 187 YDIERTIRALGDKQLTAVAASYGTQIALYAEFRPSSIRSIVFDGVVD--IDDLNDNFSW 244
OY 270 ---QWAGQOTAFNDYAAOCARSPACPLRTGDSQWNRNHALVDPLVQKRGKSDPRGISTY 326
DB 245 KLRQASHVQTEFFDRPAAMCARTKSPCLSSDRQALHOFHQLLKHSNPLDTSRGESIST 304
OY 327 ADATGTGINALYSPPRMKYTLTSGLLGRLGSDAGDLVLADDDYDGRDDAGHYSNDQAFN 386
DB 305 DDLISLTTELLMRSSWPLTAAVAVQFSGIGYSNTEIFALNSIASF-----KYSDALG 358
OY 387 AVRCYVATPDAAPAAVADQR-----IROVAPFLSYGQFTGSAPRDLCALMPV---P 436
DB 359 VILCY---DQSEQ---LSOEGRKSRKKALADAFPAVNEFEREQSDLP-ERCELMPIHRDLO 412
OY 437 ATSTPHPAAPAGAGVYVYVSTHDPATPRQSVDLARQLGAPLITFDGTQHTAVFDG-NQ 495
DB 413 QTRLNNTVLPSC---LLFVAHKYDPTPIWARKMADEFSAPLITINDGHIALAGTNL 469

```

QY	436	CYDSAVMHYFL	506
DB	470	CAVEAVERHLL	480
RESULT 4			
	PIP_AERSO	STANDARD:	PRT: 424 AA.
AC	P46547:		
DT	01-NOV-1995 (Rel. 32, Created)		
DT	01-OCT-1996 (Rel. 34, Last sequence update)		
DT	01-OCT-1996 (Rel. 34, Last annotation update)		
DE	PROLINE IMINOPEPTIDASE (EC 3.4.11.5) (PROLYL AMINOPEPTIDASE).		
GN	PAP.		
OS	Aeromonas sobria.		
OC	Bacteria; Proteobacteria; gamma subdivision; Aeromonadaceae;		
OC	Aeromonas.		
RN	[1]		
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.		
RX	MEDLINE-95189781; PubMed-7883756;		
RA	Kitazono A., Kitano A., Tsuru D., Yoshimoto T.;		
RT	"Isolation and characterization of the prolyl aminopeptidase gene (pap) from Aeromonas sobria: comparison with the Bacillus coagulans enzyme.";		
RL	J. Biochem. 116:818-825(1994).		
CC	-I- FUNCTION: HIGHER ACTIVITY TOWARD LONG PEPTIDES. ACTS ON		
CC	HYDROXYPROLINE BETA-NAPHTHYLAMIDE WITH ALMOST AS HIGH AN ACTIVITY		
CC	AS ON PROLINE BETA-NAPHTHYLAMIDE.		
CC	-I- CATALYTIC ACTIVITY: RELEASE OF A N-TERMINAL PROLINE FROM A		
CC	PEPTIDE.		
CC	-I- SUBUNIT: HOMOTETRAMER.		
CC	-I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S33.		
CC			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/or_send_an_email_to_license@isb-sib.ch).		
CC			
DR	EMBL: D30714; BAA06380.1; -		
DR	MROPS: S33.001; -		
DR	INTERPRO: IPR000073; -		
DR	INTERPRO: IPR002410; -		
DR	PFAM: PF00561; abhydrolase.1.		
DR	PRINTS: PR00793; PROAMNOPTRASE.		
KR	Hydrolase; Aminopeptidase.		
FT	INIT MET 0		
FT	ACT_SITE 145 145	POTENTIAL.	
SO	SEQUENCE 424 AA: 48274 MW, ESDD968EA90047DC CRC64;		
Query Match 4.7%; Score 128.5; DB 1; Length 424;			
Best Local Similarity 22.3%; Pred. No. 0.047;			
Matches 85; Conservative 41; Mismatches 143; Indels 113; Gaps 17;			
QY	76	SVYVDYDQPGGTQAKLAVIRVPATGQRFGA---	LLVNPGRGASAVDVAAMAPAIADT 131
	:	:	:
DB	19	TVPLDHOQPDDEETITLFGRLCRRKRDIDDELFWLLYDGGFG---	FGAPRPSANGG 72
QY	132	DI--LHNFHDLVGPRGVGSHSTPALRORTDAEPFAYRDRPMADVSSPAGYVHVEDYVQQL 188	
	:	:	:
DB	73	WIKRALDERVYLLDQRGHSHSP-----	IHEL 101
QY	189	AQDCVDRMGFSFSLATIGTASVARDMDVROALGDDQ--INVLGYSTGKLGTAULFEFGTH 247	
	:	:	:
DB	102	LAHLNPRQQAQADVLISFRADSTIVROAEILREDSRPHNSLGLGSGSGCSLTYLSLFPDS 161	
QY	248	VRAWVLDGAIDRPAVSPLESTISQM--AGFQTAENDYAADCARSP-----ACPLSTD 296	
	:	:	:
DB	162	LHEVLTG---GAVPISRSADDEVYRATYORVADKRNRAFFARFPAQAIANRLATVHQRN 217	

```

Oy      297 SNAWNRHRLADPLVQKPKGKTSRDGRSGYADATGTINATXSPORMYKTISGLLGORG   356
        :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      218 DVRLPNNGRILVEQLQQ-----QGIDL--GASGAEEELX-----YL-----    251
Oy      357 SDAGDLLVLADDYDGRDAADGHYSNDODA---FNA-----VRCYAAPPPADAAMVAAD   406
        :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      252 -----LEDFAFIETKLNPFAFLYOVOAMPNTNPFPAIHLEIHYCEGAA--SHMAA--    299
Oy      407 ORIROVPAPLISYGQ-----FTMG 423
        :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      300 ERVRGEFPALMAGCKDFAFITG 321

RESULT      5
COBG_PSEDE ID COBG_PSEDE STANDARD: PRT: 459 AA.
AC          p21637;
DT          01-MAY-1991 (Rel. 18, Created)
            01-FEB-1996 (Rel. 33, Last sequence update)
            01-FEB-1996 (Rel. 33, Last annotation update)
            COBG PROTEIN (EC 1.-.-.-).
GN          COBG.
OS          Pseudomonas denitrificans.
OC          Bacteria; Proteobacteria.
RN          [1]
RP          SEQUENCE FROM N.A.
RX          STRAIN SC510;
           MEDLINE=91008976; PubMed=2211521;
RA          Crouzet J., Cameron B., Cauchois L., Rigault S., Rouyez M.-C.,
           Blanche F., Thibaut D., Debussche L.;
RT          "Genetic and sequence analysis of an 8.7-kilobase Pseudomonas
           denitrificans fragment carrying eight genes involved in
           transformation of precorrin-2 to cobyrinic acid."
RL          J. Bacteriol. 172:5980-5990(1990).
CC          -I- FUNCTION: CATALYZES THE ELIMINATION OF C-20 IN PRECORRIN-6W TO
           FORM PRECORRIN-6X.
CC          -I- PATHWAY: COBALAMIN BIOSYNTHESIS.
CC          -I- SIMILARITY: THE C-TERMINAL DOMAIN IS A 4FE-4S/SIROHEME DOMAIN
           FOUND IN NITRATE REDUCTASES (EC 1.6.6.4 AND EC 1.7.7.1) AND
           SULFITE REDUCTASES (EC 1.8.1.2 AND EC 1.8.7.1).
CC          -----
Cc          This SWISS-PROT entry is copyright. It is produced through a collaboration
Cc          between the Swiss Institute of Bioinformatics and the EMBL Outstation --
Cc          the European Bioinformatics Institute. There are no restrictions on
Cc          use by non-profit institutions as long as its content is in no way
Cc          modified and this statement is not removed. Usage by and for commercial
Cc          entities requires a license agreement (See http://www.isb-sib.ch/announce/
Cc          or send an email to license@isb-sib.ch).
Cc          -----
DR      EMBL, M59301, AAA25795.1; -.
DR      PIR, B36145; B36145.
DR      INTERPRO: IPR000660; -.
DR      INTERPRO: IPR001899; -.
DR      PRINTS: PR00397; SIROHAEM.
DR      PROSITE: PS00365; NTR_SIR: 1.
KW      Cobalamin biosynthesis; Porphyrin biosynthesis; Oxidoreductase; Heme;
KW      Iron sulfur.
KM      METAL              338       IRON-SULFUR (4FE-4S) (POTENTIAL).
FM      METAL              344       IRON-SULFUR (4FE-4S) (POTENTIAL).
FT      METAL              377       IRON-SULFUR (4FE-4S) (POTENTIAL).
FT      METAL              381       IRON-SULFUR (4FE-4S) AND SIROHEMA
                                   (BY SIMILARITY)).
SO      SEQUENCE         459 AA; 46690 MW; EC101F4CIFFDC7209 CRC64;

Query Match      4.1%; Score 113.5; DB 1; Length 459;
Best Local Similarity 23.2%; Pred. No. 0.59; Mismatches 152; Indels 93; Gaps 20;
Matches 89; Conservative 49;

Oy      22 GLLLAAPLAGATEEPGAGQTGGAPVAVAQDSWNSCREFIADTSEIRTARCATVSVPDY 81
        ||| | |||| | | | | | | | | | | | | | | | | | | | | | | | |
Db      106 GLAIPEVPLAGIDIDEIPRIA-----TELREAL-DVRGVPLKLPKLSIVIDS 155

```

```

OY 82 DDPGTOFKLAVIRPAPGQRFQ-ALLVPPGPG--ASAVQVMVAMAPADDTILRHFD 138
Db 156 GGRFELGAVVADIRLQAVSTVAGVAMVYSLGTSKRSKASVGLTACNAVVPALITLLEKLA 215
OY 139 LVGFDPGCG---VGHSTPALRCRTDAEFDAYRRDMADYSPAGYTHVBOVYROLAQCVD 194
Db 216 SLGTTMRGRDLDPSERIALRCRETSS-----RPAAPRSAA---IGIHALGNADTVL 265
OY 135 RMGSEFLANIGTASVARMMDVROLGDDQINYL-GVSYGKRLGATLYERGTIVRAVYL 253
Db 266 GIGLAF-AOVERAALASTLHOV-QALGANALRLAPGHAF-----FVL 305
OY 254 DCAIDPAVSPTEIESIQMAGFQTAEND---YAADCARPACPLGTDSAQWVNRHYALVDP 310
Db 306 -GLCPETAANAAGSLASHGFRIAQDPRNMIATCAGSKGC-----ASAMMET-KGMER 357
OY 311 LVQ-----KPGK---TSDRGLSTYADATGTINAL---YSP 340
Db 358 LVETAPELLDSLVYHLTSGACGAKCARPKPSELTIVGAPSG--YGLVVGANGPLSAYTD 415
OY 341 QRMKYLTSGLG--LQRGSDAGD 361
Db 416 ENMGSLALRIGRLVQRKNKAGE 438

RESULT 6
PSF_HUMAN
ID PSF_HUMAN STANDARD: PRT: 707 AA.
AC P23246;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE PTB-ASSOCIATED SPLICING FACTOR (PSF).
GN PSF.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiinae; Homo.
RN [1]
RP TISSUE=FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=FETAL BRAIN;
RX MEDLINE=93194059; PubMed=8449401;
RA Patton J.G., Porro E.B., Galceran J., Tempst P., Nadal-Ginard B.;
RT "Cloning and characterization of PSF, a novel pre-mRNA splicing
RT factor.";
RL Genes Dev. 7:393-406(1993).
RN [2]
RP SEQUENCE OF 312-707 FROM N.A.
RC TISSUE=FETAL SKELETAL MUSCLE;
RX MEDLINE=90091812; PubMed=2480877;
RA Gover H.U., Moore S.E., Dickson G., Elsom V.L., Nayak R., Walsh F.S.;
RT "Cloning and characterization of a myoblast cell surface antigen
RT defined by 24.1b5 monoclonal antibody.";
RL Development 105:723-731(1989).
CC -1- FUNCTION: ESSENTIAL PRE-mRNA SPLICING FACTOR REQUIRED EARLY IN
CC SPLICOSOME FORMATION. BINDS TO THE MAMMALIAN POLYPRIMIDINE
CC TRACTS. FORMS A COMPLEX WITH THE POLYPRIMIDINE TRACT-BINDING
CC PROTEIN (PTB).
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- ALTERNATIVE PRODUCTS: TWO FORMS OF PSF ARE LIKELY TO BE PRODUCED
CC BY ALTERNATIVE SPLICING OF THE SAME GENE.
CC -1- SIMILARITY: CONTAINS 2 RNA RECOGNITION MOTIFS (RNM).
CC -1- CAUTION: WAS ORIGINALLY (REF.2) THOUGHT TO BE MYOBLAST CELL
CC SURFACE ANTIGEN 24.1b5 AND A POSSIBLE MEMBRANE-BOUND PROTEIN
CC ECTOKINASE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```

CC -----
DR EMBL: X70944; CAA50283.1; -
DR EMBL: X16850; CAA34747.1; -
DR PIR: A43557; A43557.
DR HSSP: P19339; 1SKL.
DR INTERPRO: IPR000504; -
DR PFAM: PF00076; Trm; 2.
DR PROSITE: PS0102; RRM; 2.
DR PROSITE: PS00030; RRM_RNP_1; 1.
KM Nuclear protein; RNA-binding; mRNA splicing; Alternative splicing.
FT DOMAIN 289 304 RNA-BINDING (RNP2).
FT DOMAIN 332 339 RNA-BINDING (RNP1).
FT REPEAT 9 27 3 X 3 AA REPEATS OF R-G-G.
FT REPEAT 19 11 1.
FT REPEAT 25 21 2.
FT REPEAT 25 27 3.
FT DOMAIN 10 266 GLN/GLU/PRO-RICH.
FT DOMAIN 10 15 POLY-GLY.
FT DOMAIN 20 27 POLY-GLY.
FT DOMAIN 65 65 POLY-PRO.
FT DOMAIN 571 574 POLY-ARG.
FT DOMAIN 613 616 POLY-GLY.
FT DOMAIN 635 641 POLY-GLY.
FT VARSPLIC 663 707 PNRKPRF -> VRMIDVG (IN SHORT ISOFORM).
SQ SEQUENCE 707 AA; 76149 MW; 608D5EA95E235847 CRC64;

```

```

Query Match 3.9%; Score 105.5; DB 1; Length 707;
Best Local Similarity 20.0%; Pred. No. 3.7;
Matches 85; Conservative 36; Mismatches 139; Indels 165; Gaps 18;

```

```

QY 26 AAPPLAGATEEPFGAGOT-----GAPPVAPAQOSMNSREFIADTSEIRTRCAT 74
DB 134 APPPTSGAPPGSGPPTTPPPRAVTSAPPGAPPPPPSS----- 172
QY 75 VSVPVYDQPGTGQAKLAVIRPATGQRF-----ALLV----- 108
DB 173 -GVPTTPQAGGPRPPRAVPRGPGRPKQPGPGKGMGPKRGGGPGLSTGSHR 231
QY 109 -----NPGPGASAVDMVAAMADALADTDLIRHFDLV 140
DB 232 KPRNRGSGPRRGROHNRPRYHOONHOGPRPGGSGRSEKISDSEGFKANLSILRR--- 287
QY 141 GPRDPRGVGSTRALRGR-----PDAERDAYRDMADYSPAGVTHNVOYRQLAQDCV 193
DB 288 -----PGKTYTORCRLVGNLRADITEDEERK-LFAKYGERGEVYINK----- 330
QY 194 DMGSEFLANIGTASVARDMDVROALDDQINVLGYSTKLGTAFLRFGTHVRAVYL 253
DB 331 -GKGFEFI-KLESRALA-----ELAKAELDDTPMR-----GRDLRRFAHNAALSV 375
QY 254 DGAIDPAVSP--IEESISQMAAGFQTA-----FNDYAADCARSPA--CPLG 294
DB 376 RN-LSPYVSNELLEAFSGFPIERAVVIVDDRGSRGKGIYEFASKPARKAFERCSBG 434
QY 295 TDSAQVNNVYHALVPLVQKPKTSRPLSTAD-----ATTGTTINALYSQ 341
DB 435 VELLTTTPR-PIVIEPLLEDLDDDEGLPERKLAOKNPMYOKERETPRFAOHGTFEYYS-Q 492
QY 342 RKKYL 346
DB 493 RKKSL 497

```

```

RESULT 7
KAS2_STRCM STANDARD; PRT; 402 AA.
ID KAS2_STRCM

```

```

AC P41176;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last annotation update)
DE PUTATIVE POLYKETIDE BETA-KETOACYL SYNTHASE 2 (BC 2.3.1.-) (ORF 2).
OS Streptomyces cinnamonensis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A3823.5;
RX MEDLINE=92374994; PubMed=1508151;
RA Arrowmith T.J., Malpartida F., Sherman D.H., Birch A., Hopwood D.A.,
RA Robinson J.A.;
RT "Characterisation of acyl-homologous DNA encoding polyketide synthase
RT genes from the monensin producer Streptomyces cinnamonensis.";
RL Mol. Gen. Genet. 234:254-264(1992).
CC -1- PATHWAY: BIOSYNTHESIS OF THE POLYKETIDE ANTIBIOTIC MONENSIN.
CC -1- MISCELLANEOUS: THIS PUTATIVE KETOACYL SYNTHASE LACKS THE ACTIVE
CC SITE CYSTEINE.
CC -1- SIMILARITY: BELONGS TO THE BETA-KETOACYL-ACP SYNTHASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb.sib.ch).
CC -----
DR EMBL: Z11511; CAA77597.1; -
DR PIR: S25077; S25077.
DR INTERPRO: IPR000794; -
DR PFAM: PF00109; ketoacyl-synt; 1.
DR Antibiotic biosynthesis; Transferase; Acyltransferase.
SQ SEQUENCE 402 AA; 41270 MW; 348657D10201C214 CRC64;

```

```

Query Match 3.8%; Score 104; DB 1; Length 402;
Best Local Similarity 24.0%; Pred. No. 2.4;
Matches 115; Conservative 50; Mismatches 189; Indels 126; Gaps 29;

```

```

QY 22 GLLIAPPLAGATEEPFGAGOTGAPPVAPAQOSMNSREFIADTSEIRTRCATVSVPVYD 81
DB 8 GMGIAAPNGLG---RPTTPRPPAP-----RAASASTRF----- 39
QY 82 DQPGTGQAKLAVIRPATGQRFGLLVNPG--PGASAVDMVAAMP--AIADTDLIRHF 137
DB 40 -DPSGVPAQLAG-ELPG-----FRAAEHLRGLVPRQDRVTRLSLAADALADACV----- 89
QY 138 DLVGFPRKGVGSTRALRGRTPDAERDAYRDMADYSPAGVTHNVOYRQLAQDCVDRMG 197
DB 90 EVAAEFPLDMGVYT-----ASHAGGEFGODELQKLGGG-OPLYSAVXOSFA----- 135
QY 198 FSLFANIGTASVARDMDVROALDDQINVLGYSTKLGTA-FLRFGTHVRAVYLDGA 256
DB 136 WFLVAVNSQISTRHGKKGPGSVVVSQAGL-----DALAQARRLVKGT---PLIVCGA 187
QY 257 IDPAVSPIESISQMAFGQTAENDYAADCARSPACPLGDSQMW-NRYHALVPLVQPK 315
DB 188 VERPSRAGAGSPSPAG---GMSD--SDEPNRAYLPRFDGSGGYVGGRGVYVPL--ER 240
QY 316 GRTSDPRGLS-YADATTTGTTINALYSPOKRYLTSGLGLQKSSDAGDILLVLDYDVG--- 371
DB 241 AEAAPARGAEVYGEA--GFLARLPAH-----SG-----RSTRNAHRTALDDAGTAP 287
QY 372 -----RDADGHNSNDQAFNAVRCVYATPADRAAMVAADQRIQOVAPFLSYQGF-TG 423
DB 288 GDIRRVFADGGGRYPNDRAEAIAISEVGR-----GRVYVTCPRMTGGLSG 335
QY 424 SAPRDL-CALMPVPATSTPNPAAPAGAKVVVVSTTH-DPATPYQSGVLAQOLGAPILT 481
DB 336 AAPPLVACALL-----AMRAG-----VTPRPNHIDPRCEYLDLVLVYQVRAALRT 381

```

```

RESULT 8
ID IE18_PRIVIF STANDARD; PRT; 1461 AA.
AC P11675;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE IMMEDIATE-EARLY PROTEIN IE180.
GN IE.
OS Pseudorabies virus (strain Indiana-Funkhauser / Becker) (PRV).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicelloviruses.
RN [1]
RX MEDLINE-93315207; PubMed-2546124;
RA Cheung A.K.;
RT "DNA nucleotide sequence analysis of the immediate-early gene of
   Pseudorabies virus.";
   Nucleic Acids Res. 17:4637-4646(1989).
   [2]
REVISIONS.
RA Cheung A.K.;
RL Submitted (NOV-1989) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: THIS IE PROTEIN IS A MULTIFUNCTIONAL PROTEIN CAPABLE
   OF MIGRATING TO THE NUCLEUS, BINDING TO DNA, TRANS-ACTIVATING
   OTHER VIRAL GENES, AND AUTOREGULATING ITS OWN SYNTHESIS.
CC -1- SUBCELLULAR LOCATION: NUCLEUS OF INFECTED CELLS.
CC -1- PTM: A LONG STRETCH OF SERINE RESIDUES MAY BE A MAJOR SITE OF
   PHOSPHORYLATION.
CC -1- SIMILARITY: BELONGS TO THE ICP4/IE140/IE180 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X15120; CA33214.1; -.
DR PIR: S04713; EDBE1F.
DR KM DNA-binding; Phosphorylation; Nuclear protein.
FT DOMAIN 390 405 POLY-SER.
FT FT 958 966 POLY-SER.
FT SEQUENCE 1461 AA; 149833 MW; 7F31E7ABE403B208 CRC64;

Query Match 3.8%; Score 103.5; DB 1; Length 1461;
Best Local Similarity 21.4%; Pred. No. 13;
Matches 121; Conservative 54; Mismatches 200; Indels 191; Gaps 31;

OY 26 AAPPLAGATEEPGAGGPGAPVAPQGSWNSCR-EFTADTSEIFTRACATVSPVDYDQ 84
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
DB 172 AASPASSSSSGSGSGSPG-PSAPRR-WSPARQDVPGEPP--AARPRTPAPPA---QP 224
OY 85 GGTGAKLAVIRVPTGGRFGLLVNPGGPGASAVDMVAAMPAIADDDILRHFP----- 138
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
DB 225 A-----AVAAAPR--RKG--ASPAAPGAPVAPGSGGSGDGRGHHQHREPL 273
OY 139 -----LVGDPFGVGHSTPALRCRTAEFDYRRDPMADYSPAGVTHVEQVYRQLADC 192
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
DB 274 LDPEAARRIDPRRLGARSPV-----SSNPSSSSSTTYTAAVEVAVAGPERKD- 320
OY 193 VDRMGESFLANIGTASYARDMDVRAQLDQDQVY-LGYSTYKLGTAVERFGTHVRAM 251
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
DB 321 EDGIG--LAGDGGAPLQROP--RRRAGEGALRRGRGFSSSSGSGS----- 362
OY 252 VLDSADIDPANS-----IEEISQMAQFQTAINDVAAACARSPACPLG 294
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
DB 363 --DSDLSPANSPAPRAPAAAAAAAAAARSASSSSSSSSSSSSSSSEGEDEGVRRGAPL- 419

```

```

OY 295 TDSAQWNRVHALVDPVQKPKGTS DPR-GLSYADATGTINA-----LSPQRMKY 345
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
DB 420 -----ARAGPPSPPPAPAAAPRPSASSATSSSSAAAPAPAPAPAPRRPKRR 468
OY 346 LTSGLLGLR-----GSD-----AGDLLV-LADDYDGRDA 374
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
DB 469 STNNHLSIMADGPEPTDPLTLPLGEPWPGSDPADGRVRYGAGDSREGIMDEVDVRA 528
OY 375 DGHY-----SNDQDAFNA-VRCVYAPTPADPAWAAQRIQVAPFLSYG 419
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
DB 529 AARYRAAGPVPEPIPMGSRKQHEALVRLTYSAGAGEAMSWL---QNPMAQPDGRFN 585
OY 420 QFTGSAPRDLCAIWPVATSTPHPAAPAGAKVYVSTT-----HDP--ATPY 465
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
DB 586 QF-----CQRVH--APHGHSFTTGSVTPPLPHIGDMAAQPDLMLPH 628
OY 466 Q-SCVDLAROGLAPLTFDGTQHTAV 490
DB 629 AVSAVAMSR-----YDRTQKTFI 647

RESULT 9
MCAS_MYCBO STANDARD; PRT; 2110 AA.
ID MCAS_MYCBO
AC Q02251;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE MYCOCEKOSIC ACID SYNTHASE.
GN MAS.
OS Mycobacterium bovis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterinae; Mycobacteriaceae; Mycobacterium.
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE-92406887; PubMed-1527058;
RA Mathur M., Kolattukudy P.E.;
RT "Molecular cloning and sequencing of the gene for mycocecosic acid
   synthase, a novel fatty acid elongating multifunctional enzyme, from
   Mycobacterium tuberculosis var. bovis Bacillus Calmette-Guerin.";
   J. Biol. Chem. 267:19388-19395(1992).
RL J. Biol. Chem. 267:19388-19395(1992).
CC -1- FUNCTION: CATALYZES THE ELONGATION OF N-FATTY ACYL-COA WITH
   METHYLMALONYL-COA (NOT MALONYL-COA) AS THE ELONGATING AGENT TO
   FORM MYCOCEKOSYL LIPIDS.
CC -1- COFACTOR: CONTAINS 1 COVALENTLY BOUND PHOSPHOPANTHETHEINE.
CC -1- SUBUNIT: HOMODIMER WHOSE MONOMERS PROBABLY HAVE A HEAD TO TAIL
   ARRANGEMENT.
CC -1- SUBCELLULAR LOCATION: MEMBRANE-ASSOCIATED.
CC -1- SIMILARITY: PARTIAL TO S.ERYTHRAEA ERYTHRONOLIDE SYNTHASE. MODULE
CC 4, AND TO VERTEBRATE FATTY ACID SYNTHASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M95808; AAA25369.1; -.
DR PIR: B44110; B44110.
DR HSSP: Q02054; IAF8.
DR INTERPRO: IPR000255; -.
DR INTERPRO: IPR000794; -.
DR INTERPRO: IPR001227; -.
DR INTERPRO: IPR002085; -.
DR PRAM: PF00698; Acyl_transf_1.
DR PRAM: PF00107; adh_zinc_1.
DR PRAM: PF00109; ketoacyl-synt_1.
DR PRAM: PF00550; pp-binding_1.
DR PROSITE: PS00012; PHOSPHOPANTHETHEINE, 1.
DR PROSITE: PS00606; B_KETOACYL_SYNTHASE, 1.

```

```

CC      -!- FUNCTION: REQUIRED FOR THE ACTION OF COLICINS K AND L AND FOR THE
CC      STABILIZATION OF MATING AGGREGATES IN CONJUGATION. SERVES AS A
CC      RECEPTOR FOR A NUMBER OF T-EVEN LIKE PHAGES. ALSO ACTS AS A PORIN
CC      WITH LOW PERMEABILITY THAT ALLOWS SLOW PENETRATION OF SMALL
CC      SOLUTIONS (BY SIMILARITY).
CC      -!- SUBUNIT: MONOMER (PROBABLE).
CC      -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.
CC      -!- SIMILARITY: BELONGS TO THE OMPA FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; X00618; CAA25254.1; -.
DR      PIR; S07298; S07298.
DR      INTERPRO; IPR000498; -.
DR      INTERPRO; IPR001145; -.
DR      INTERPRO; IPR002368; -.
DR      PFAM; PF00691; OmpA; 1.
DR      PFAM; PF01389; OmpA_membrane; 1.
DR      PRINTS; PRO1021; OMPADOMAIN.
DR      PRINTS; PRO1022; OUTRAMBRANE.
DR      PROSITE; PS01068; OMPA; 1.
KW      Outer membrane; Transmembrane; Phage recognition; Signal; Repeat;
KW      Porin.
FT      SIGNAL          1       21
FT      CHAIN           22      359  OUTER MEMBRANE PROTEIN A.
FT      TRANSMEM        27      40  POTENTIAL.
FT      TRANSMEM        62      74  POTENTIAL.
FT      TRANSMEM        77      92  POTENTIAL.
FT      TRANSMEM       104     114  POTENTIAL.
FT      TRANSMEM       118     133  POTENTIAL.
FT      TRANSMEM       154     165  POTENTIAL.
FT      TRANSMEM       171     187  POTENTIAL.
FT      TRANSMEM       193     204  POTENTIAL.
FT      DOMAIN         210     219  5 X 2 AA TANDEM REPEATS OF A-P.
FT      REPEAT         210     211  1.
FT      REPEAT         212     213  2.
FT      REPEAT         214     215  3.
FT      REPEAT         216     217  4.
FT      REPEAT         218     219  5.
FT      DOMAIN         268     312  OMPA-LIKE.
FT      DISULFID       322      336  BY SIMILARITY.
SQ      SEQUENCE       359 AA; 38426 MW; 13992A03C19758B CRC64;
Query Match          3.7%; Score 101; DB 1; Length 359;
Best Local Similarity 22.6%; Pred. No. 3.3;
Matches 65; Conservative 27; Mismatches 91; Indels 104; Gaps 12.
QY      95 RVPATGQFFGALLVPGGPGASAVDMVAAMADAIADTDILRHFDLVGEDPRGVGHSTPAL 154
DB      88 RMPYKGS-----VNNGAFFKAQGVQLAKLISYPIAD-DLDIYTRILGGM----- 128
QY      155 RCRDAEFDAYRRDMADY-----SPAGYTHVQVYRQLAQCDCVDRKMGPSF 200
DB      129 -----VWRADSKANRGRTGRLSDHDGVSPLAAVGYEA---LTKNATRIIDYQF 176
QY      201 LANIGTASVARDMDVROLGDGDDINTLGYSYGTGLGTAYLERFGGTHVRAMVLDAIDPA 260
DB      177 VSNISDACTV-----GARPDNTMLSLGVS-----RFGQ-----DDVYAPA 212
QY      261 VSP-----IEESISQMAQFQTAFNDAADCARSP-----ACPLGTD 296
DB      213 PAPAPAPVYETKRTFLKSPDLVLFNFKSKTLAKGQQAALDQLYQLGSLMPPKDGSSVVLGYT 272
QY      297 SAQWVNRH-----ALVDPLQVK--PGKTSPPRGSLSTADATG 332
DB      273 DAVGSDOYNQKLSQORAQSVVDYLVSKGIPSKISARGNGEADAVTG 319

```

```

RESULT 11
AMYA_ASPAW STANDARD; PRT; 498 AA.
AC 002905;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 01-OCT-1996 (Rel. 34, Last annotation update)
DE ALPHA-AMYLASE A PRECURSOR (EC 3.2.1.1) (1,4-ALPHA-D-GLUCAN
DE GLUCANOHYDROLASE A).
GN AMYA.
OS Aspergillus awamori.
OC Eukaryota; Fungi; Ascomycota; Eurotiales; Trichocomaceae;
OC anamorphic Trichocomaceae; Aspergillus.
RN [1]
RC SEQUENCE FROM N.A.
RX STRAIN-UVK143F;
MEDLINE-90254827; PubMed-2340591;
Korman D.R., Bayliss F.T., Barnett C.C., Carmona C.L.,
Kodama K.H., Royer T.J., Thompson S.A., Ward M., Wilson L.J.,
Berka R.M.:
"Cloning, characterization, and expression of two alpha-amylase genes
from Aspergillus niger var. awamori."
Curr. Genet. 17:203-212(1990).
CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-ALPHA-GLUCOSIDIC
CC LINKAGES IN OLIGOSACCHARIDES AND POLYSACCHARIDES.
CC -1- COFACTOR: BINDS A CALCIUM ION REQUIRED FOR ITS ACTIVITY.
CC -1- SIMILARITY: BELONGS TO FAMILY 13 OF GLYCOSYL HYDROLASES, ALSO
CC KNOWN AS THE ALPHA-AMYLASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X52755; CAA36966.1; -
CC DR HSSP: P10528; 6TAA.
CC DR INTERPRO: IPR000461; -
CC DR PFAM: PF00128; alpha-amylase; 1.
CC KW Hydrolase; Glycosidase; Carbohydrate metabolism; Calcium;
CC Glycoprotein; Signal; Multigene family.
CC FT SIGNAL 1 21 POTENTIAL.
CC FT CHAIN 22 498 ALPHA-AMYLASE A.
CC FT ACT_SITE 227 227 BY SIMILARITY.
CC FT ACT_SITE 231 231 BY SIMILARITY.
CC FT ACT_SITE 251 251 BY SIMILARITY.
CC FT ACT_SITE 318 318 BY SIMILARITY.
CC FT DISULFID 51 59 BY SIMILARITY.
CC FT DISULFID 171 185 BY SIMILARITY.
CC FT DISULFID 261 304 BY SIMILARITY.
CC FT DISULFID 461 496 BY SIMILARITY.
CC FT CARBOHYD 218 218 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC SEQUENCE 498 AA; 54880 MW; 7658511BC01A8A01 CRC64;
Query Match 3.7%; Score 101; DB 1; Length 498;
Best Local Similarity 19.4%; Pred. No. 5;
Matches 118; Conservative 64; Mismatches 183; Indels 242; Gaps 33;

```

```

DB 101 HGYWODIYSLNENYGTADDLKALSSALHERGMYLMWVYVANNHGYD--GAGSS----- 152
QY 157 RTAAEFDAYRRDPMADXPAGVTHVEQYROL--AODCDMAGFSFLAN-----I 204
DB 153 -----VDIS-----VFKPFSSQDFH--PFCFIONEDQTOVEDCWL 187
QY 205 GTASVA-RDMDMVROALGDDOINYLG-----YSY-GTKLCT-----AYLERGTH 247
DB 188 GDMTVSLPDDTTRKDYVKNEMVWVGLSVNSYSLDGLRIDTVKHVQKDFPGRNKAAGVY 247
QY 248 VRAMVLDGALDPA-----VSPIESISOMAGFOAFNDYAADCA 286
DB 248 CIGEVLDG--DPAVTCYQVNMVGVLMYPITYPLCLMFKSTGSMDDLYNMINTVSDCP 305
QY 287 RSPACPFGTDSADQWNVNHYHALVPLVQPKGKTSPPRLSYADATYTGINALYSPQRKYL 346
DB 306 DSLT--LGT-----FVEHN-----DNPRASTN----- 327
QY 347 TSGLLGLQSGSDAGDLLVLADD-----YGRADAGHYSNDDAFNVCYVAPPADPAW 402
DB 328 -----DIALKNVAAPFILNDGIPITVYAGQEQ--HVAAGNDPANREATWLSGVPTSELY 380
QY 403 --VAADRIQVA-----PLSTGQFTGSAPRDLCLMPPATSTHPAPACAGKYVYV 455
DB 381 KLASRNAIRNVAISKDTGEVTKN-----WPIYKDDTITPMKRGTDGSOIVT 428
QY 456 STYHDPAT-----PYOSGVDLAROLGAPLTFDQGTHTAVPDGQCVDSAVMHY 504
DB 429 ILSKKGASGSYTLISLGAQITAGQOLTEYIGCTTVYGS-----DGNVPVPM----- 477
QY 505 FLDGTLF 511
DB 478 ---GGLP 481
RESULT 12
AMYA_ASPAW STANDARD; PRT; 499 AA.
AC 002906;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE ALPHA-AMYLASE B PRECURSOR (EC 3.2.1.1) (1,4-ALPHA-D-GLUCAN
DE GLUCANOHYDROLASE B).
GN AMYA.
OS Aspergillus awamori.
OC Eukaryota; Fungi; Ascomycota; Eurotiales; Trichocomaceae;
OC anamorphic Trichocomaceae; Aspergillus.
RN [1]
RC SEQUENCE FROM N.A.
RX STRAIN-UVK143F;
MEDLINE-90254827; PubMed-2340591;
Korman D.R., Bayliss F.T., Barnett C.C., Carmona C.L.,
Kodama K.H., Royer T.J., Thompson S.A., Ward M., Wilson L.J.,
Berka R.M.:
"Cloning, characterization, and expression of two alpha-amylase genes
from Aspergillus niger var. awamori."
Curr. Genet. 17:203-212(1990).
CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-ALPHA-GLUCOSIDIC
CC LINKAGES IN OLIGOSACCHARIDES AND POLYSACCHARIDES.
CC -1- COFACTOR: BINDS A CALCIUM ION REQUIRED FOR ITS ACTIVITY.
CC -1- SIMILARITY: BELONGS TO FAMILY 13 OF GLYCOSYL HYDROLASES, ALSO
CC KNOWN AS THE ALPHA-AMYLASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X52756; CAA36967.1; -

```


DR HSP: P10529; 6TAA.
 DR INTERPRO: IPR000461; .
 DR PFAM: PF00128; alpha-amylase; 1.
 KM Hydrolyase; Glycosidase; Carbohydrate metabolism; Calcium;
 KW Glycoprotein; Signal; Multigene family.
 FT SIGNAL 1 21
 FT CHAIN 1 21
 FT ACT_SITE 227 499
 FT ACT_SITE 227 499
 FT ACT_SITE 231 231
 FT ACT_SITE 251 251
 FT ACT_SITE 318 318
 FT ACT_SITE 318 318
 FT DISULFID 51 59
 FT DISULFID 171 185
 FT DISULFID 261 304
 FT DISULFID 461 496
 FT CARBOHYD 218 218
 SQ SEQUENCE 499 AA; 54921 MW; 740B96B1BC01A8A CRC64; N-LINKED (GLCNAC...) (POTENTIAL).

Query Match 3.7%; Score 101; DB 1; Length 499;
 Best Local Similarity 19.4%; Pred. No. 5;
 Matches 118; Conservative 64; Mismatches 183; Indels 242; Gaps 33;

QY 16 FGLLGLLAAPPLAGATEEPGAGOTPGAPVVAPOQSNNS-----CREFIADTSEIRT 69
 DB 6 WSLFLYGLGLOVAPALAA-----TPA-----DMRSOSITFLLDREFARTDGSSTT 48
 QY 70 ARCATVSVPVYDOPGQTGA---KLAVIRVPATGQRFGLLVNPGS---PGASA----- 117
 DB 49 ARCNT-----ADQKVCSTGMOGIDKLDYIO---GMGFATWITPTAQLPQTAVGDAY 100
 QY 118 -----VQVMAAPALADTLRHFDLV---GDPFGVGHSTALRC 156
 DB 101 HGYMODIYSLNENYGTADLKLSSALHERGMYLWVDVYANMGYD--GAGSS----- 152
 QY 157 RTDAEPDARVDPMDYSPAGVTHVEQYRQL-AQDCVDMGSPFLAN-----I 204
 DB 153 -----VDYS-----VKRPSSQDYFH--PCFTQNTEDQTOVEDCWL 187
 QY 205 GTASVA-RDMWVROALGDDQINYLQ---YSY-GTKLGT-----AYLERFGTH 247
 DB 188 GDMTVSLPDLDTKDYVKNEMVDMGSLVSNYSIDGLRIDTVKHVQKDFWPGYNKAAGYV 247
 QY 248 VRAMVLDGALDPA-----VSPIESISOMAGFOTAFNDYADCA 286
 DB 248 CIGEVLDG--DPAYTCPYQNVMDGVNLNPIYPLLNFRKSTSGSMODLYNMITVKSDCP 305
 QY 287 RSPACPLGTDSAQWNRVYHALVDPLVOKPGKTSDPRLGSYADATGTGINALYSPORMKYL 346
 DB 306 DSTL--LGT-----EVENH-----DNPRFASTTN----- 327
 QY 347 TSGLLGLQKSGDAGDLLVLAD--YDGRDAGHYSNDOAFNAVRCVYAPTPADPAAM 402
 DB 328 -----DIALAKNVAFLILNDGIPILYAGEQ--HYAGNDPAMRATWLSGPTDSELY 380
 QY 403 --VAAQORIROVA-----PFLSYGFTGSAPRDLCAIMPATSTHPPAPAGAGVYV 455
 DB 381 KLLASNAIRNVAISKDTGTVYKN-----WPIYKDDTIIIPMRKGTSGQSIYV 428
 QY 456 STTHDPT-----PYOSGVDLAROLGAPLITPDTQTOHTAVFQNOCVDSVVMYH 504
 DB 429 ILSNRKASGDSYTLISLSGAGYTAGQOLTEYIGCTTVGVS-----DGNVPVPM-- 477
 QY 505 FLDGTLR 511
 DB 478 ---GGLP 481
 RESULT 13
 LYS9_YEAST STANDARD; PRT; 446 AA.
 AC P3899;
 DT 01-FEB-1995 (Rel. 31, Created)

DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE SACHAROPINE DEHYDROGENASE [NADP+, L-GLUTAMATE FORMING] (EC 1.5.1.10).
 GN LYS9 OR LYS13 OR YNR050C OR N3461.
 OS Saccharomyces cerevisiae (Baker's yeast).
 EU Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
 OC Saccharomycetaceae; Saccharomyces.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=1278B;
 RA Feller A.;
 RL Submitted (JAN-1994) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Pohl T.M.;
 RL Submitted (MAY-1996) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 314-324.
 RC STRAIN=ATCC 38531 / Y41;
 RX MEDLINE=97089742; PubMed=8935650;
 RA Norbeck J., Blomberg A.;
 RT "protein expression during exponential growth in 0.7 M NaCl medium of
 RT Saccharomyces cerevisiae";
 RL FEMS Microbiol. Lett. 137:1-8(1996).
 CC -I- CATALYTIC ACTIVITY: N6-(L-1,3-DICARBOXYPROPYL)-L-LYSINE + NADP(+) + H(2)O = L-GLUTAMATE + 2 AMINOADIPATE 6-SEMIALDEHYDE + NADPH.
 CC -I- PATHWAY: SEVENTH STEP IN LYSINE BIOSYNTHESIS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC EMBL: X77363; CAA54552.1; -
 DR EMBL: Z71665; CAA86331.1; -
 DR PIR: S41937; S41937.
 DR YEPD: 8416; -
 DR SGD: S0005333; LYS9.
 KW Lysine biosynthesis; Oxidoreductase; NADP.
 SQ SEQUENCE 446 AA; 48917 MW; 1EC4CE1CE3BCD916 CRC64;

Query Match 3.6%; Score 99.5; DB 1; Length 446;
 Best Local Similarity 20.9%; Pred. No. 5.6;
 Matches 80; Conservative 37; Mismatches 125; Indels 141; Gaps 17;

QY 107 LVNPGGPGASAVDMV--AAMAPAIADTDLIRHFDLVGFDPFGVGHSTPALRCTDAEFDA 164
 DB 42 LAKPSSKKAISLDVTDSDALDKVLADNDVYLSLIPYTFPHNV--KSAIRKRTDVTSS 98
 QY 165 YRRDPADYSP---AGVT-----HVEQYRQLADQCVDRMG---FSPLANIGTAS 208
 DB 99 YISPALRELEPELVKAGITVYMEIGLDPGIDHLYAVKTIDEVHRAKGAKLSFLSTGGP 158
 QY 209 VARDMDVROALGDDQINLYGYSYGTGKLTAVYLERGTHVRAVL-----DGAID 258
 DB 159 APEDSD-----NPLGY-----KRSWSRGVLLALRNSAKYKRGKIE 195
 QY 259 PAVSPIEESISQMGAGQTAENFYADACASPACPLCTDSAGQVNRH----- 305
 DB 196 YVSS---EDLMATAKPYFTYPGAFA---VCYPRNDSTLEFDLYHDEAETVINGTLR 246
 QY 306 -----ALVDPVLVOKPGKTSDPRLGSYADATGTGINALYS--PQRMKYLTSGLGLQR 355
 DB 247 YGCFPEFVKAIVD-----MGMLKDDA-----NEIFSKPIAMNEALKQYIG--AK 288
 QY 356 GSDAGDLVLVADYDGRDADGHSNDQDAFNNAVRCVYAPTPADPAAMVAAQORIROVAPF 415
 DB 289 STSKEDLIASID-----SKATWQDEDEHERITLSGF 318

DR EMBL: 274410; CAA98940.1; -
DR TUBERCULIST: RV0107C; -
DR INTERPRO: IPR000695; -
DR INTERPRO: IPR001757; -
DR PFAM: PF00122; E1-E2_Atpase; 1.
DR PRINTS: PR00119; CATATPASE.
DR PRINTS: PR00120; CATATPASE.
DR PROSITE: PS00154; ATPASE_E1_E2; 1.
KW Hydrolyase; Transmembrane; Phosphorylation; ATP-binding.
FT TRANSMEM 30 50 POTENTIAL.
FT TRANSMEM 148 168 POTENTIAL.
FT TRANSMEM 177 197 POTENTIAL.
FT TRANSMEM 315 335 POTENTIAL.
FT TRANSMEM 358 378 POTENTIAL.
FT TRANSMEM 637 657 POTENTIAL.
FT TRANSMEM 673 693 POTENTIAL.
FT TRANSMEM 778 798 POTENTIAL.
FT TRANSMEM 921 941 POTENTIAL.
FT TRANSMEM 969 989 POTENTIAL.
FT TRANSMEM 997 1017 POTENTIAL.
FT TRANSMEM 1401 1421 POTENTIAL.
FT TRANSMEM 1432 1452 POTENTIAL.
FT TRANSMEM 1547 1567 POTENTIAL.
FT MOD_RES 1053 1053 PHOSPHORYLATION (BY SIMILARITY).
SQ SEQUENCE 1632 AA; 169606 MW; A29E651A55EF7FD8 CRC64;

Query Match 3.6%; Score 99; DB 1; Length 1632;
Best Local Similarity 22.1%; Pred. No. 30;
Matches 109; Conservative 57; Mismatches 235; Indels 92; Gaps 20;

OY 23 LLAAAPLAGATEEPGAGTGPAPVYAPPOOSMNSCREFTADTSEIRT----- 69
DB 1116 LLAAASLAGSLSSQGDSEW---VLAEPFESSRGYAAIGRGTDGIPMLMKAPE 1171
OY 70 ---ARCATVSVPYDQPGTQAKLAV--IRVPATGOR-EGALLVNPGGASAVDMVAA 123
DB 1172 TLPRCLADPGVDHHAESVVRHLAEOGLRYLAVAOFTWNGTTHDETDADAVDAVA- 1230
OY 124 MAPAIDTDILRHFDLVGDFPRGVGHSTPALRCRTDAEFDAYRRDPADYSPAGVTIVEQ 183
DB 1231 ---HDLLEIGY--VGLADTARSSSRPLIELLDAB-----RNVVLITGDHPIT-ARA 1276
OY 184 VYRQLADQCDVRM---GFSFLANIGTASYARMDMVRQALGDDQINLYGISTYKLGTA 239
DB 1277 IARQLGLPADARVVTGAEELAVLDEEHAHKLADMQVFARVSPKQVOIV-----A 1326
DB 240 YLERFSTHRAWYLDGAIDPAVSPIEESTISQMGFQTAFNDAADCARSPACPLGTDSSAQ 299
DB 1327 ALORCG-RVTAMVGDAADAAA-----IRNADVGIGVSGRGSSAARGAADIIVLTDDDL 1378
OY 300 WNNRYHALVDPYQKPGKTSDFR-----GLSYADATGTGINALYSPQRMKYLTSGLLG 352
DB 1379 GV-----LIDALVEGSRMAGVADATILVGGVGEVLEFTVIGTAFAGARAPVGTROLL 1433
OY 353 LORGSAGDLVLA--DDYGRDAGHYSDQDAFNAVR-----CYAAPADPAANVAAD 406
DB 1434 VNLITDMPALAVAVTSQPAEPDDAEYPTDDAERAQREHRAVLLIGTPPSIDA----- 1487
OY 407 ORIROVAPFLSYGQFTGSAAPROLCALMPV--ATSTPHPAAPGAGKVVVSTTHDPATP 464
DB 1488 PLIRQI---VNRGVTAAG---ATAAMAIGRKTPTGERTATWGLTALVMTQLAQTLTLR 1541
OY 465 YOSGVDIAROLGA 477
DB 1542 RNSPLVIATALGS 1554

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 5, 2001, 14:28:28 ; Search time 67.42 Seconds
(without alignments)
522.700 Million cell updates/sec

Title: US-09-461-774-4

Perfect score: 2738

Sequence: 1 AAMWRRRLSSALLSFGILL.....AVMHYFLDGTLPPTSLRCAP 519

Scoring table:

BLOSUM62
Gap 10.0 , Gapext 0.5

Searched: 195891 seqs, 67900655 residues

Actual number of hits satisfying chosen parameters: 195891

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2717	99.2	520	2	C70776
2	1286.5	47.0	520	2	D70776
3	780	28.5	539	2	T36660
4	583.5	21.3	529	2	T35966
5	547.5	20.0	543	2	T34681
6	529	19.3	528	2	T35352
7	298.5	10.9	528	2	D70968
8	271	9.9	226	2	B36944
9	261	9.5	526	2	G82572
10	128.5	4.7	425	2	JC4184
11	119	4.3	1079	2	A70972
12	116.5	4.3	585	2	E70930
13	115	4.2	447	2	T34992
14	114	4.2	1400	2	B70963
15	113.5	4.1	459	2	B36145
16	112.5	4.1	274	2	A56716
17	112.5	4.1	1815	2	E73021
18	112.5	4.1	7576	2	T17428
19	112	4.1	643	2	E70682
20	111	4.1	2591	2	T30288
21	110.5	4.0	2082	2	T37056
22	110	4.0	995	2	C83203
23	109.5	4.0	671	2	F82960
24	109	4.0	725	2	T35114
25	109	4.0	1055	2	H70951
26	108.5	4.0	383	2	H75263
27	108.5	4.0	574	2	F75356
28	108.5	4.0	751	2	T46517
29	108	3.9	525	2	E75374

30	108	3.9	588	2	T35549	hypothetical prote
31	108	3.9	815	2	T00538	serine proteinase
32	107	3.9	485	2	T51109	probable reductase
33	107	3.9	1639	2	T14181	peptide synthetase
34	106.5	3.9	515	2	T47946	prolyl aminopeptid
35	106.5	3.9	2314	2	T28698	hypothetical prote
36	106.5	3.9	4613	2	T17409	polyketide synthas
37	106	3.9	409	2	A70932	probable ppe prote
38	106	3.9	571	2	T36317	two-component sens
39	105.5	3.9	452	2	T35762	probable amino aci
40	105.5	3.9	552	2	T29114	probable secreted
41	105.5	3.9	629	2	E47096	membrane transloca
42	105.5	3.9	707	2	A46302	PTB-associated spl
43	105	3.8	673	2	B70895	hypothetical prote
44	105	3.8	2468	2	A83412	hypothetical prote
45	105	3.8	2569	2	T14164	peptide synthetase

ALIGNMENTS

Query Match	Score	2717	DB 2	Length	520
Best Local Similarity	99.4%	Pred. No. 2.8e-184			
Matches	516	Conservative	1	Mismatches	2
				Indels	0
				Gaps	0
QY	1	AAMWRRRLSSALLSFGILLGGLLAAPPLAGATEEPGAGCPVVAPOOSMNSCRFF	60		
DB	2	AAMWRRRLSSALLSFGILLGGLLAAPPLAGATEEPGAGCPVVAPOOSMNSCRFF	61		
QY	61	IADTSEIRFARCATVSVPVYDQPGTQAKLAVIVPATGQFEGALLVNPFGASAVDM	120		
DB	62	IADTSEIRFARCATVSVPVYDQPGTQAKLAVIVPATGQFEGALLVNPFGASAVDM	121		
QY	121	VAAAPALADTDILRHFDLVGFDPGVGHSTPALRCRTDAERDAYRDPADYSPAGVTH	180		
DB	122	VAAAPALADTDILRHFDLVGFDPGVGHSTPALRCRTDAERDAYRDPADYSPAGVTH	181		
QY	181	VQOVYROLAQDCVDMGSEFLANIGTASVARDMDVROLGSDQINYLGYSTGLGTAY	240		
DB	182	VQOVYROLAQDCVDMGSEFLANIGTASVARDMDVROLGSDQINYLGYSTGLGTAY	241		
QY	241	LERFSTHRAVYLDGALIDPAVSPVIESISQMGFQTAENFYAADCARSPACPLGTDSAQW	300		
DB	242	LERFSTHRAVYLDGALIDPAVSPVIESISQMGFQTAENFYAADCARSPACPLGTDSAQW	301		
QY	301	VNRHYALVDPLVQKPGKTSDPRLGSLYADATTGTINALYSPOKRYLTSGLLQKGSAG	360		
DB	302	VNRHYALVDPLVQKPGKTSDPRLGSLYADATTGTINALYSPOKRYLTSGLLQKGSAG	361		
QY	361	DLVLVADVDGADGHSNDQDAFNAVCYAFIPADPAAMVAVADQRIQVAPFLTSQ	420		

Db 362 DLVLYADDDYGRADGHHXNDQDAFNNAVRCVDAPTPADPAAWAAOQRIROVAPELSTYQ 421
|||||
QY 421 FTGSAPRDLCALMPVATSTPHPAAPAGAKVYVSTTHPAPTPYOSGVDLARQLAPLI 480
|||||
Db 422 FTGSAPRDLCALMPVATSTPHPAAPAGAKVYVSTTHPAPTPYOSGVDLARQLAPLI 481
|||||
QY 481 TFDGTOHTAVFDGNOCVDSAVMHYFLDGLTLPPTSLRCAP 519
|||||
Db 482 TFDGTOHTAVFDGNOCVDSAVMHYFLDGLTLPPTSLRCAP 520
|||||

RESULT 2
D70776
Probable export proteinase Rv2224c - Mycobacterium tuberculosis (strain H37Rv)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: D70776
R:Cole, S.T.; Broesch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.;
andream, M.A.; Davies, R.; Devlin, K.; Felwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
Mature 393, 537-544, 1998
A:Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A>Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; M0ID:98295987
A:Accession: D70776
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-520 <COL>
A:Cross-references: GB:270692; GB:AL123456; MID:93261567; PIDN:CAA94647.1; PID:91261926
A:Experimental source: strain H37Rv
C:Genetics:
A:Gene: Rv2224c

Query Match 47.0%; Score 1286.5; DB 2; Length 520;
Best Local Similarity 48.8%; Pred. No. 2.5e-83;
Matches 254; Conservative 69; Mismatches 183; Indels 15; Gaps 5;

QY 5 RRRPLSSALLSFGLLGILLAPPLAGATEEPGAGOTPGAPVYAPQOSNNSCREFIADT 64
|||
Db 7 RRRKIRMLIMALAAVALV---LVGCIYVGGRRMAEPKIGORPVENTPCRSSNPQY 62
|||
QY 65 SEIRTRCATVSVVDYDQGTQAKLAVIRVPATGQRFGLLVNPGSPASAVDMAAM 124
|||
Db 63 KIRGALCGKLANPVYDRDGDVALALALIRFATGDKIGSLVINPGPESGIEALGV 122
|||
QY 125 APAIADTILRHFDLVGDFPRGVGHSTPALRCRTDAEFDAYRRDPMAVSPAGVTHVEY 184
|||
Db 123 FQTLPRR-VHERFDLVGFDRGVASSRPATWCNSDADNDRLRAEPQVYDSREGVAHLENE 181
|||
QY 185 YRLADQCVRMGEFSLNIGTASVARMDMVRQALGDDQINLYGYSYGRKLTAVERF 244
|||
Db 182 TKQFVRCVDMKKNFLAHGTVNAVDALRAALGDDQITLYLGSYGRISATAIEEF 241
|||
QY 245 GTHVRAVMDLGAIDPAVPSPIESISQWAGFOTAFNDYAADCAAPACPLCTDSANOWNRY 304
|||
Db 242 PQGVRAVMDLGAIDPAVPSPIESISQWAGFOTAFNDYAADCAAPACPLCTDSANOWNRY 301
|||
QY 305 HALVDELV-----OKPGTSDPRGLSTADATGTTINALYSQRMKYLTLGSLGLOGRS 357
|||
Db 302 HSLVDELVDPNDRISRPATKDPKGLSYSDAIVGTIMALYSPNLMOHLLDGLSELV--D 359
|||
QY 358 DAGD-LLVYADDDYGRADGHHXNDQDAFNNAVRCVDAPTPADPAAWAAOQRIROVAPEL 416
|||
Db 360 NRQDTLLALADMMRDSDHRTNNSGDARVAINCVDOPVTDKQYIDERRARETAPFPM 419
|||
QY 417 SYGQFTGSAAPRDLCALMPVATSTPHPAAPAGAKVYVSTTHPAPTPYOSGVDLARQL 476
|||
Db 420 SYGQFTGSAAPRDLCALMPVATSTPHPAAPAGAKVYVSTTHPAPTPYOSGVDLARQL 479
|||
QY 477 APLTFDGTQHTAVFDGNOCVDSAVMHYFLDGLTLPPTSLRC 517
|||

Db 480 GSLLTFDGTQHTAVFDGNOCVDSAVMHYFLDGLTLPPTSPGAKC 520
|||||
RESULT 3
T36660
proteinnase - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T36660
R:Seeger, K.J.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M
submitted to the EMBL Data Library, March 1999
A:Accession: T36660
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-539 <SEE>
A:Cross-references: EMBL:AL035636; PIDN:CA838476.1; GSPDB:GN00070; SCSCH5.02c
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: slpD; SCSCH5.02c

Query Match 28.5%; Score 780; DB 2; Length 539;
Best Local Similarity 37.2%; Pred. No. 1.5e-47;
Matches 203; Conservative 69; Mismatches 220; Indels 54; Gaps 19;

QY 4 WRRRPLSSALLSFGLLGILLAPPLAGATEEPGAGOTPGAPVYAPQO-----52
|||
Db 17 FRATLTLAALLATACSGAGSTAGSPAKAKAGATEATATTLTP-LPKATPAELSPYEQ 75
|||
QY 53 --SWNSCREFIADTSEIRTRCATVSVVDYDQGTQAKLAVIRVPAT--GQRFGLLV 108
|||
Db 76 KLGMRDC-----GVPRGQCATMKAPLDYAKAPADGVDLAVARRKATGFCRLSLIV 127
|||
QY 109 NPGSPGASAVDMAAMAPALADTILRHFDLVGDFPRGVGHSTPALRCRTDAEFDAYRRD 168
|||
Db 128 NPGSPGASAVDMAAMAPALADTILRHFDLVGDFPRGVGHSTPALRCRTDAEFDAYRRD 185
|||
QY 169 PMADYSP--AGVT--HVEQYVROLAODC-VDRMGFSFLANTGTSVARMDMVRQALGDDQ 224
|||
Db 186 --TDVTPDDAGETDELVDAYKEFEGCGADAP--KLIRHSTVEAARDMDVLAVALGLK 241
|||
QY 225 INLYGYSYGRKLTAVERFTHVRAVMDLGAIDPAVPSPIESISQWAGFOTAFNDYAAD 284
|||
Db 242 LTVYGASYGFTGATYAGLEPRDGRVLVDGAMPPLPARLMLDQEGEYAFQGFAXD 301
|||
QY 285 CAAPSPACPLG---TDSAQWYRNHALVDELVQK--FGKTSDPKGLSYADATGTTINALYS 339
|||
Db 302 CVKQPCPCPLGDKDTPDQVGNLKSFEEDLDAPPLPAGADGRLKTESLATTTGVIAAMYD 361
|||
QY 340 PQRMKYLTLGSLGLOGSPDAGDGLLVYADDDYGRADGHHXNDQDAFNNAVRCVAPTPADP 399
|||
Db 362 EGAMQOLRESLTAIEKDGAGALLISDSYEREADGSGYNLMFAAAVNCIDL-----P 416
|||
QY 400 AAHVAAADQRIQVAPPLS-----YQFTGSAPRDL-CALMPVATSTPHPAAPAGAKVY 453
|||
Db 417 AAFSPDE-VRDALDPFEKASPVFGE--GLAWSSLNCAYVVPKPTGEPHRIEAGATPIV 473
|||
QY 454 VVSTHDPATPYOSGVDLARQL-GAPLITFDGTQHTAVFDGNOCVDSAVMHYFLDGLTLP 512
|||
Db 474 VVSTHDPATPYOSGVDLARQL-GAPLITFDGTQHTAVFDGNOCVDSAVMHYFLDGLTLP 512
|||
QY 513 TSLRCA 518
|||
Db 534 DGRKCS 539
|||

RESULT 4
T35966
probable secreted peptidase - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T35966

Db 65 EMPBDLQCGVTVPLDLSRGRGTLVDALAKYRATGDSKSVLLNFGPGSSGINELALG 124
125 APAIADIDLRHFDLVGDFRGVGHSTR-----ALRCRTDAEFDAIRPDMADYSPAGV 178
125 GKER -MHILINGDYVTFDFRGVGRSSPVGCRATLKIMETDDBGGMTDPEG----- 175
179 THVEQVYROLAQDCVDRMGESFLANIGTASVARDMDVNRQALGDDQINVLGYSTKLG 238
176 -VLERLRDAAAECAKYSG-PVLPHTIGTVDAARDMDVNRRLGDDRLNLYGFSYGTBLGA 232
239 AYLEFRTGTHRAWYLDALIDPAVSPT-EESISQMAFGOTAPNDYAAACASPCPLGTDS 297
233 VYARAFPDKGRWVLDG-VPTLTERPLAEQGLAGARGOQTALLENLDMCVEDVACPFQODA 291
298 AQWNRHVALVDPLVYKPGKTSYADATGTI-----NALYSPORMKVL----- 346
292 RDARDQYERLVASL-----DSDPVSAGFEPPTGQDMVCAIGCALYSRELMSLEBALA 345
347 -----TSGILGLQSGSDAGDLVLADVDYDGRDAGHYSNDOD-----AFNAVRCV 391
346 QLEEDGTRGLEGGSSGVTFEPVRAAVNAPVRGDRTAGLDEEDVPMDNLPAALMALINCA 405
392 YATPPADPAAMVAADQIR-----QVAP-----FLSYGQFTGS-----APRICAL 432
406 DDDPRTAAQVTSALDRARAREDEVSPFGRYRLTQVLMCGYGRPGTDYTRDDYKDL--- 462
433 WVPYATSTPBAAPAGAKGVVSTTHDPATPYOSGVDLAROLCAPLITFD--GTOHTAV 490
463 -----DTRARMLVGTBGDPATPTKWTETETADRLGPSAVVLNDNGEGHTG- 506
491 FDGNQCVDSAVMHYFLDGLPTSLRCAP 519
507 YASSCKVHRKVDLFXGLSPDQSSGCP 535

RESULT 7

hypothetical protein RV2672 - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: D70968
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feldwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrett, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; M0ID:98295987
Accession: D70968
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-528 <COL>
A:Cross-references: GB:Z80225; GB:AL123456; NID:g3242265; PIDN:CAB02326.1; PID:e266414;
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: RV2672

Query Match 10.9%; Score 298.5; DB 2; Length 528;
Best Local Similarity 25.1%; Pred. No. 1.4e-13;
Matches 142; Conservative 72; Mismatches 250; Indels 101; Gaps 26;

QY 7 RPLSSALLSRGLLGGLLAA-----PPLAGATPEPCAGQTPGAPV 48
DB 9 RPTTSTRAMLVLCSATVLAACVPAGCADPRATYSGAGQGAATTTTPPAGP---PPLA 65
QY 49 APOQ--SWNSCREFIADTSEIRTA-----RCATVSVVDVDQDGGTQAKAIVRPA--T 99
DB 66 APRNDLSMHCTSRVYSNAGIPAPGVKLECASTYDDLDLVGGSTAVSIGVRAANSOT 125
QY 100 GQRFGALLVNPGGASAVDMVAMAPALADTDLRHFDLVGDFDPGRVGHSTPALRCRTD 159

Db 126 PSDAGLIVETGSDLPSTOLPVWLAH--AGIDILRSHPIVAVDRMGMSPP-IDCR-- 180
QY 160 AEPDAYRRDPMADYS--PAG-----VTHSEQVYROLAQDCVDRM--GESFLANIGTASVARD 212
Db 181 ---DHFRDDEMROAQORQADDDPYANLSDLSNTATTDCTALAPGESAYNTHTAAS---D 234
QY 213 MDWYROLGDDQINVLGYSTYKLGTAYLEFRTGTHRAWYLDALIDPAVSPTLEBSISQMA 272
Db 235 IERLRKLMVDPALAFVIGTGQVALYAAASRPDNRARLILDSPIALGVSAEAAAEQOVQ 294
QY 273 GFQTAFFNDVADCARSPACPLGSDSNQWNRHVALVPLVYKPGKTSYADATGTI 332
Db 295 GQQAALDAFAAQCV-AVNCALGSHP-----KGAVSALLSAARSGDGPASVAAVANA 346
QY 333 TINALYSPQRMKY-LTSGILGLQSGSDAGDLVL-----ADDYDGRDAGHY--SNDQDA 384
Db 347 VATRALGPPDSGRDSTTKLADALAAASGMNLLSALINRAD--TTDPTGQFTSSGSDA 404
QY 385 FNAVRCVYATPPADPAAMVAADQIRQ--VAPFLSYGQFTGSAPRDL--CALMPVPATST 440
Db 405 VN-----RPTP-----DRVELVYAMGKLYPQFGAVALNLVKCVHM--PSSSP 446
QY 441 PHPAAPGACKVYVSTTHDPATPYOSGVDLAROLCAPLITFDGTOHTAVFDG----- 493
Db 447 PGPCKDLV-DVLLGLQNDPIV---GNEGVAATAATATNANAASKRYMVGIGHCASI 501
QY 494 -NOCVDSAVMHYFLDGLPTSLRC 517
Db 502 YSSCAVPLVAYLDTGKLPDPTDYTC 526

RESULT 8

B36944
hypothetical protein (comp 3' region) - Escherichia coli (fragment)
C:Species: Escherichia coli
C:Date: 07-Jul-1995 #sequence_revision 07-Jul-1995 #text_change 19-Oct-1995
C:Accession: B36944; S37474
R:Kautmann, A.; Stierhof, Y.D.; Henning, U.
J. Bacteriol. 176, 359-367, 1994
A:Title: New outer membrane-associated protease of Escherichia coli K-12.
A:Reference number: A36944; M0ID:94117370
Accession: B36944
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-226 <KAU>
A:Cross-references: GB:X74278
C:Genetics:
A:Gene: ompP
C:Keywords: membrane protein

Query Match 9.9%; Score 271; DB 2; Length 226;
Best Local Similarity 29.1%; Pred. No. 4e-12;
Matches 64; Conservative 43; Mismatches 79; Indels 34; Gaps 6;

QY 47 VVAPOQSWNSC-----REFIADTSEIRTAFCATVSVVDVDQDGGTQA-----KLAIV 94
Db 25 VSARQIEMQSCMTSPYSDWGKSSSEFLCGYLSVPLKTTDYGKVDSENIPLVRLAMP 84
QY 95 RVPATGQRFGALLVNPGGASAVDMVAMAPAI---ADTDLRHFDLVGDFDPGRVGH 150
Db 85 KLPAKSKRKSVILLISGGPLPGIN-----PYINDMWPTNLRSDWIDTIGDFPRGVGS 138
QY 151 TPAIRCTDAEFDAYRRDPMADYSVAGYTHVEQVYROLAQDCVDRMGESFLANIGTASVA 210
Db 139 FPAINCQOS-----MOERLVNVE-----KQILLQKINACIHNHGAELVIRHIGSEAV 186
QY 211 RDMWYROLGDDQINVLGYSTYKLGTAYLEFRTGTHRAWYLDALIDPAVSPTLEBSISQMA 250
Db 187 YDIERIRQALGDKQTLAVAYSTGTQIALYAEKFPSSIRS 226

RESULT 9

G82572
 proteinase Xf2330 (imported) - Xylella fastidiosa (strain 9a5c)
 C:Species: Xylella fastidiosa
 C:Date: 18-Aug-2000 #sequence-revision 20-Aug-2000 #text-change 20-Aug-2000
 C:Accession: G82572
 R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequencing
 Nature 406, 151-157, 2000
 A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
 A:Reference number: A82515; MUID:20365717
 A:Note: For a complete list of authors see reference number A59328 below
 A:Accession: G82572
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-526 <SIM>
 A:Cross-References: GB:AE004043; GB:AE003849; NID:g9107486; PIDN:AA65129.1; GSPDB:GN001
 A:Experimental source: strain 9a5c
 R:Simposon, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
 Britones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carrer, D.M.; Carreir, H
 as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.
 Submitted to GenBank, June 2000
 Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm
 , D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laizy
 chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Martino, C.L.; Marques, M.V.; Martins, E
 A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
 , F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmeri, D.A
 Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santeili, R.V.; Sawasak
 A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
 M.; Tsuchiko, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
 A:Reference number: A59328
 A:Contents: annotation
 C:Genetics:
 A:Gene: XF2330

Query Match 9.5%; Score 261; DB 2; Length 526;
 Best Local Similarity 25.3%; Pred. No. 6, 1e-11;
 Matches 140; Conservative 76; Mismatches 217; Indels 120; Gaps 30;
 7 RPLSSALLSFGILLGULLAAPPLAGATEPFGAGTPGAPVAP--QGSNSCRREFIADT 64
 13 RMLKKYIVLALVGVGFVFPTRVAASSFASAAATLQPVNSYGRLEPEPILSARGA 72
 65 SEIRTRCATVSPVVDYDGGTQAKLAVIRV-----PATGQRGALLVNGGEGASA 117
 73 GHRVL-CQQLPVLNPALEGRITLKIAMLERDSCGHPASPVF---FVAGGEGQSA 127
 118 VDMVAAMAPADTDILRHFDLVGDFPRGVGSHSTPALRCRTDAEFDAYRRDPMAVSPAG 177
 128 TGVVAAVAPSL--NEVRRKKRDIFFVDRGTGSHR--LTC--VDAQRAFOLE-----AE 176
 178 VTHVEOVYRQLADQCDVDM--GFSFLANIGTASVARDMDVRLAGDDQINVLGYSTGKL 236
 177 NTTPPEQYIAFVORCAASLOGRADPRYTTTEAVTDLDMVRLASVDKINLLGVSXGTRY 236
 237 GTATLERFEGTHVAMVLDGIDPAVSPIEEST-SQMA-GPOTAFNFIYADCAASPC--- 291
 237 AOHYARIRAHVFTLMDGVY-----PNDLVIGSEFATTFENALADQCKRDPSCAKR 291
 292 -PLGTSAGMVNRVHALVPLVQKPGKTSDPRLSYADATTGTINALYSPQRKYLTSGL 350
 292 FPL--DSRQGLR-----TIMERLRQSGP--VEYQDPASGEM-----RRRIRIADVY 333
 351 LGLQGR-SPAGDLVLAADYDGDADGHS-----NDQAFN-----AVRCVYA 393
 334 VSLAFGLSYAPELIALPLILNEASGHYAPRLMSLSQVLGMDMDQMHALQAVICT-- 391
 394 PPADAAVAVADQRTQVAPFLSYGQFTGSAPRDL-----CALMP-----VATGTS 440
 392 ---EDADRIVAS-----ANAGAGALFGESEVAAQMFPPCKKWPGRIRPDPFWMVLSVD 441
 441 PHPAAPAGAGKVVVSTTHDPATP-----YOSGVDLAROLGAPLITFGTOHFAFDEN 494
 442 ----LPA-----LLSGEDRPVTPPHYAKLKLGLRYGHIYAP-----GQGGITIRVG- 486

QY 495 QCVDSAVMHYFLD 507
 DB 487 -CV-SKLLGQFID 497

RESULT 10

JC4184
 prolyl aminopeptidase (EC 3.4.11.5) - Aeromonas sobria
 M:Alternate names: Pro-X aminopeptidase; proline aminopeptidase; proline iminopeptida
 C:Species: Aeromonas sobria
 C:Date: 31-Aug-1995 #sequence-revision 27-Oct-1995 #text-change 20-Jun-2000
 C:Accession: JC4184; PC4051; I39543
 R:Kitazono, A.; Kitano, A.; Tsuru, D.; Yoshimoto, T.
 J. Biochem. 116, 818-825, 1995
 A:Title: Isolation and characterization of the prolyl aminopeptidase gene (pap) from
 A:Reference number: JC4184
 A:Accession: JC4184
 A:Molecule type: DNA
 A:Residues: 1-425 <KIT>
 A:Cross-References: DDBJ:D30714; NID:9604877; PIDN:BA06380.1; PID 1236731
 A:Accession: PC4051
 A:Molecule type: protein
 R:Kitazono, A.; Kitano, A.; Tsuru, D.; Yoshimoto, T.
 J. Biochem. 116, 818-825, 1994
 A:Title: Isolation and characterization of the prolyl aminopeptidase gene (pap) from
 A:Reference number: I39543; MUID:95189781
 A:Accession: I39543
 A:Status: preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: DNA
 A:Residues: 1-385 <RES>
 A:Cross-References: GB:D30714; NID:9604877; PID:9604878
 C:Comment: This enzyme is widely distributed in nature, it catalyzes the removal of a
 e activities.
 C:Genetics:
 A:Gene: pap
 C:Superfamily: prolyl aminopeptidase
 C:Keywords: alpha-aminoacylpeptide hydrolase
 F:2-425/Product: prolyl aminopeptidase #status predicted <MAT>

Query Match 4.7%; Score 128.5; DB 2; Length 425;
 Best Local Similarity 22.3%; Pred. No. 0.1;
 Matches 85; Conservative 41; Mismatches 143; Indels 113; Gaps 17;
 76 SVFVDYDQPGTQAKLAVIRVATGQFGA---LLVNGGEGASAVDMVAMAPAIADT 131
 20 TVPLDHOQDDDETITLFGRTLCRKRLDDELPMILYLOGGR-----FGAPRPSANG 73
 132 DI--LRHFDLVGDFPRGVGSHSTPALRCRTDAEFDAYRRDPMAVSPAGVTHVEOVYRQL 188
 74 WIKRALQEFRRVILLDDRGCHSTR-----THAEL 102
 189 ADDCVDRMGFSFLANIGTASVARDMDVRLAGDDQ--INVLGYSTGKTGTAYLEFGTH 247
 103 LAHLNPRQADVLISHRADSVIRDAELREQLSPDPHWSI;LQSGFSGSLYLSLFPDS 162
 248 VRAMVLDGADIDPAVSPIEESISOM--AGFOTAFNDVAAACASP-----ACPLGTD 296
 163 LHEVYITG---GVATIGSADEVYRATYQRAYADKKRAFAFPNAAQALANRLATLQNH 218
 297 SAQMVNRVHALVPLVQKPGKTSDPRLSYADATTGTINALYSPQRKYLTSGLGLQGR 356
 219 DVRLPGRGRLTVEQLQO-----GIDL--GASGAFRELY-----YL----- 252
 357 SPAGDLVLAADYDGDADGHSNDQDA---FNA-----VRQVYAPRPADPAVAVAD 406
 253 -----LEDAFTGKRLNPAFLYQVAAQPFNTNPFVAILHELHYCEGAA--SHWA-- 300
 407 ORLRQVAPFLSYGQ-----FTG 423
 301 ERYRGEPALANAGKDFRFTG 322

Tue Jan 16 10:17:18 2001

us-09-461-774-4.rpr

Page 8

```

Db      306  --GACPEETAAVAAGSLAASHGRINEDOPRNALATCAAGSKGC-----ASAMMET-KGAER 357
Qy      311  LVQ-----KPGK--TSPRGISYADATTTGNTL-----YSP 340
Db      358  LVETPELLDQSLVYHLSGCAGKCAKREKPSBELTVGAPSG--YGLVYNGAANGLPSPAYTD 415
Qy      341  QRWKLYTSGLLG--LQKSGDAGD 361
Db      416  ENKGSALALRGLRVLROKNDAGE 438

```

Search completed: January 5, 2001, 14:28:34
Job time: 255 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 5, 2001, 14:27:16 ; Search time 45.19 Seconds
(without alignments)
206.234 Million cell updates/sec

Title: US-09-461-774-4

Perfect score: 2738
Sequence: 1 AAMWRRRLSSALLSFGILL.....AVMHYFLDGLTPRLCAP 519

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 174772 seqs, 17957048 residues

tal number of hits satisfying chosen parameters: 174772

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA:*

1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*\n3: /cgn2_6/ptodata/2/1aa/6.COMB.pep:*\n4: /cgn2_6/ptodata/2/1aa/CTUS.COMB.pep:*\n5: /cgn2_6/ptodata/2/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	780	28.5	539	1	US-08-173-508-8
2	780	28.5	539	2	US-08-265-310-8
3	780	28.5	539	3	US-08-951-742-8
4	653.5	23.9	513	1	US-08-173-508-6
5	653.5	23.9	513	2	US-08-265-310-6
6	653.5	23.9	513	3	US-08-951-742-6
7	570.5	20.8	537	1	US-08-173-508-2
8	570.5	20.8	537	2	US-08-265-310-2
9	570.5	20.8	537	3	US-08-951-742-2
10	108.5	4.0	1996	2	US-08-804-227C-9
11	108.5	4.0	1996	2	US-08-804-198-3
12	108.5	3.9	4551	3	US-09-320-878-1
13	106	3.9	571	2	US-08-796-414B-1
14	104	3.8	423	3	US-09-335-409-10
15	101	3.7	3724	2	US-08-804-227C-10
16	101	3.7	3724	2	US-08-804-198-4
17	98	3.6	913	3	US-08-911-853-17
18	96	3.5	1156	3	US-08-996-083-11
19	94.5	3.5	2756	1	US-08-375-709-11
20	94.5	3.5	2756	1	US-08-752-929-11
21	94.5	3.5	2756	3	US-09-090-793-7
22	93	3.4	1864	2	US-08-804-227C-3
23	92	3.4	1235	2	US-08-680-326-36
24	91.5	3.3	2172	1	US-08-611-107-31
25	91.5	3.3	3567	2	US-07-642-734C-4
26	91.5	3.3	3567	2	US-08-439-009A-4
27	90.5	3.3	3729	2	US-08-804-227C-4
28	89.5	3.3	3011	1	US-08-453-552-2

29	89.5	3.3	3011	2	US-08-710-637-2	Sequence 2, Appl
30	89.5	3.3	3011	4	PCT-US93-00907-2	Sequence 2, Appl
31	89.5	3.3	4828	3	US-09-036-987A-5	Sequence 5, Appl
32	89	3.3	1298	2	US-08-690-473-2	Sequence 2, Appl
33	89	3.3	5588	3	US-09-036-987A-6	Sequence 6, Appl
34	88.5	3.2	594	3	US-08-729-416C-17	Sequence 17, Appl
35	88.5	3.2	1164	2	US-08-589-756-1	Sequence 1, Appl
36	88	3.2	485	3	US-09-320-878-10	Sequence 10, Appl
37	87.5	3.2	584	1	US-08-426-819A-36	Sequence 36, Appl
38	87.5	3.2	622	1	US-08-426-819A-35	Sequence 35, Appl
39	87.5	3.2	638	2	US-08-426-125-5	Sequence 5, Appl
40	87.5	3.2	638	2	US-08-455-355-5	Sequence 5, Appl
41	87.5	3.2	3170	2	US-07-642-734C-5	Sequence 5, Appl
42	87.5	3.2	3170	3	US-08-439-009A-5	Sequence 5, Appl
43	87	3.2	3739	3	US-09-320-878-2	Sequence 2, Appl
44	86.5	3.2	1240	2	US-08-680-326-37	Sequence 37, Appl
45	86	3.1	594	3	US-08-729-416C-7	Sequence 7, Appl

ALIGNMENTS

RESULT 1
US-08-173-508-8
Sequence 8, Application US/08173508
Patent No. 5616485
GENERAL INFORMATION:
APPLICANT: Bartfield, Daniel
APPLICANT: Butler, Michael J.
APPLICANT: Hadary, Dany
APPLICANT: Jenish, David
APPLICANT: Krieger, Timothy
TITLE OF INVENTION: STREPTOMYCES PROTEASES AND IMPROVED
TITLE OF INVENTION: STREPTOMYCES STRAINS FOR EXPRESSION OF PEPTIDES AND
TITLE OF INVENTION: POLYPEPTIDES
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W.
CITY: Washington, D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/173,508
FILING DATE: 23-DEC-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET INFORMATION: 18740/125/CACO
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202 672 5300
TELEFAX: 202 672 5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 539 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-173-508-8
Query Match 28.5%; Score 780; DB 1; Length 539;
Best Local Similarity 37.2%; Pred. No. 26-64;
Matches 203; Conservative 69; Mismatches 220; Indels 54; Gaps 19;
QY 4 WRRRLSSALLSFGILL-----APPLAGATEEPGAGGTPCAPVAPQO----- 52

```

Db      17 FRATLLTAAELATACASGASSTASGSPAKAAGAATEATATLTLP-LPKAPPAELSPYEQ 75
Qy      53 ---SWMSCEFLADISEITARCATYSPVVDQGGCGTQAKAVIRVAT--GQRFGLLY 108
Db      76 KLGWDC-----GVGFQCATMKAPLADYAKPADGQVRLAARAKKATGPKKRLGSLV 127
Qy      109 NPGGASAVDMVMAAMAFAIADTILRHFDLVGFDPGVCHSTPALKCRDPAEFDVARD 168
Db      128 NPGGGGSAIGYIQYQYAGIGYPARYAQYDVAVDPGVASEP-VELDQREMDATTR- 165
Qy      169 PMADSP--AGVT-HVEQYRQLAQC-VDRMGFSFLANIGTASVARDMDVNRQALGDDQ 224
Db      186 --TDVTPDAEETELVDAYKEFEGGADAP--KLIRHVSVAARDMDVLRAYVLGDEK 241
Qy      225 INVLCSYSGTGLGATYLERFCTHVRAMWLDAIDPAVSPIEESTISOMAGQOTANDYAAD 284
Db      242 LTYGASATGTLGATYAGLPFDRFGRVLVDGAMPDPLPARLNLQEQEGETARQSKAKD 301
Qy      285 CARSPACPLG---TDSAGWNRHYHALVDPLVQ--PKGTSDDPRLSYADATTGTTINALS 339
Db      302 CVKQDPCDLGKRDPTPDQVGNKLSFFEDDLDAKPLPAGDADGRKLTSLATTGVIAAMYD 361
Qy      340 PQRMKYLISGLLGQKRGSDADDLLVLADYDGRADGHIYSNDQAFNAFVACVAPPTADP 399
Db      362 EGAMQQLSESTLSAIKEKDGAGLLILSDSYXERADGYSNLMFANAAVNCLEL-----P 416
Qy      400 AAMVAADRIKOVAPFLS-----YGFQTSAPRDL-CALMPVPTSTPHPAAPAGACKVY 453
Db      417 AAFSSPDE-VADALPDPFKASPVGE--GLAWSSLNCAVYKRTGEPHRIEAGAATPIY 473
Qy      454 VVSTTHDPATPYQSGVDLAROL-CAPLTEDGTQHTAVFDGNOCVDSAVMHYFLDGLTPP 512
Db      474 VVGTRDPATPYRMAEALSDQLTSGHLITLEGDGHATYAGRSSCIDSAINTYLLTGAPE 533
Qy      513 TSLRCA 518
Db      534 DGRKCS 539

RESULT 2
US-08-265-310-8
: Sequence 8, Application US/08265310
: Patent No. 5856166
: GENERAL INFORMATION:
:   APPLICANT: Bartfeld, Daniel
:   APPLICANT: Butler, Michael J.
:   APPLICANT: Hadary, Dany
:   APPLICANT: Jenish, David
:   APPLICANT: Krieger, Timothy
:   APPLICANT: Malek, Lawrence F.
:   APPLICANT: Soostmeyer, Gisela
:   APPLICANT: Walczyk, Eva
:   APPLICANT: Krygman, Phyllis
:   APPLICANT: Garven, Sheila
:   TITLE OF INVENTION: STREPTOMYCES PROTEASES AND IMPROVED
:   TITLE OF INVENTION: STREPTOMYCES STRAINS FOR EXPRESSION OF PEPTIDES AND
:   TITLE OF INVENTION: POLYPEPTIDES
:   NUMBER OF SEQUENCES: 21
:   CORRESPONDENCE ADDRESS:
:   ADDRESSEE: Foley & Lardner
:   STREET: 3000 K Street, N.W.
:   CITY: Washington, D.C.
:   COUNTRY: USA
:   ZIP: 20007-5109
: COMPUTER READABLE FORM:
:   MEDIUM TYPE: Floppy disk
:   COMPUTER: IBM PC compatible
:   OPERATING SYSTEM: PC-DOS/MS-DOS
:   SOFTWARE: Patentin Release #1.0, Version #1.25
:   CURRENT APPLICATION DATA:
:   APPLICATION NUMBER: US/08/265,310
:   FILING DATE: 24-JUN-1994

```

```

: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/173,508
: FILING DATE: 23-DEC-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: BENT, Stephen A.
: REGISTRATION NUMBER: 29,768
: REFERENCE/DOCKET NUMBER: 18740/13/CACO
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 202 672 5300
: TELEFAX: 202 672 5399
: TELEX: 904136
: INFORMATION FOR SEQ ID NO: 8:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 539 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-265-310-8

Query Match      28.5%; Score 780; DB 2; Length 539;
Best local similarity 37.2%; Pred. No. 2e-64;
Matches 203; Conservative 69; Mismatches 220; Indels 54; Gaps 19;

QY      4 WRRRLPSALLSFFGLLGLLA----APPLAGATEEPGACGTPGAPVAPQO----- 52
      17 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      17 FRATLLTALLATACAGAGASTSAGSPAKAKAGATEAATATLTLP-LPRATPAELSPYEQ 75

QY      53 ---SNMSCEFIADTSEIRKCAIVSVPYVDQPEGCTQAKLAVIVPAT--GQRFALLY 108
      76 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      76 KLGMWDC-----GVGPGFCATWKAPLDYAKKPADGDVRLAVAKKATGPKRGLSELY 127

QY      109 NPGCGASAVDMVAAMAPAIADTDLRHFDLVGFPDPGCVGHSTPAALRCRTDPEFAYRD 168
      128 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      128 NPGGGSAGIAGLQGYAGIGYPAKRAQYDMVAVDPRGVASEP--VECLDGHENDAYTR- 185

QY      169 PMADYSP--AGYT-HVEQYVROLAODC-VDRMGSEFLANIGTASVARDMDYKQALGDDQ 224
      186 --TDVTPDAGETDELVDAYKEFAEGCGADAP--KILRHVSTVEAARMQMDYLRAVLGDEK 241

QY      225 INVLGYSGTKIGATYLERFGHNVAMVLDAIDPAVSPIEESISOMAGFOTAFDVIAD 284
      242 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      242 LTYVASIGTFLGATYAGLFPDRTGRVLVDGAMDPSLRARLNLEQTEGFEAFQSFKD 301

QY      285 CARSPACPLG---TDSAQVMNRYNALVDPLOYK--PGKTSDEPGLSYADATTGTINALYS 339
      302 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      302 CVKPGDCPLGDKDTPDPQVGNKLKSFDDDLAKPLPAGDADGRKLTESLATGTVIAMMD 361

QY      340 PQRKTYLTSLGLLGRSGDADGLVLADYDGRDADGHYSNDQDFANVRCVYATPARDP 399
      362 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      362 EGAMQQLSESTLSAIKEKEDGALLTSSYEREDGGSNLMFMAAVNCLD-----P 416

QY      400 AAMVAADPRIRQVAPFLS-----YGFQGSAPRD-CALMPVATSPHPAPACAGKVV 453
      417 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      417 AAFSSPDE-VNDALPDEFKASPVGE--GLAMSLNCAYWPKYKPGEGHRIEAAATATIV 473

QY      454 VVSTTHDPATPYQSGVDIAROL-GAPLITPDGTOHTAVFDGNCQVDSAVMHYFLDGTLP 512
      474 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      474 VVGTRDEATPYRMAEALDQSLTSGHLLTYEGDGHYTAGRGSSCIDSAINTYLLTGAPE 533

QY      513 TSLRCA 518
      534 : : :
      534 DGRKCS 539

RESULT      3
US-08-951-742-8
: Sequence 8, Application US/08951742
: Patent No. 6127144
: GENERAL INFORMATION:
: APPLICANT: Michael J. Butler
: APPLICANT: Michael J. Butler

```


STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/951,742
FILING DATE: 16-OCT-1997
ATTORNEY/AGENT INFORMATION:
NAME: Bent, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 0189740/0140
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 513 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-951-742-6

Query Match 23.9%; Score 653.5; DB 3; Length 513;
Best Local Similarity 33.6%; Pred. No. 1.2e-52;
Matches 179; Conservative 69; Mismatches 224; Indels 61; Gaps 18;

QY 18 LLLGGLLAAPLACATEE-----PGAGTGPAPVVAPOOS--WNSCRETIADT 64
DB 10 LRAAALILTAGSGSGSDKDKDGGRSSAGPSMAAPSGPEALASQTLDMARC-----EG 64
QY 65 SEIRTA-----RCATVSVVDVDOPGGTQAKLAVIRPATG-QRGALLVNPGRGASAV 118
DB 65 SDAAPRDPGDWRKATLKAPIDMSDPGETIDLALIRSRASGDDRTGSLFNFGGPGASGV 124
QY 119 DVAAAPAPAIADT--DILRHFDLVGFDPKGVSHSTPRLKRTDAEFDAYRDMADYSPA 176
DB 125 STM-----PSYADTVSLNHERYDLVSWDPGVA-ASEGVRCRIDEALIEAESVSTPDSRA 179
QY 177 GYTHVQVYROLAOD----CVDRMGFSFLANIGTASVARDMDKVRALGDQDINTYGYST 232
DB 180 ----EQAAVLKDAADRGRCERKAG-KLMEHVSTTDRTARDMDLRRHVLDGERMHHYGCISY 234
QY 233 GYKLGTAUYLEREGTHYRAVWLDAIDPAVSPITEESISOMAGPOTAFNDYAAOCARSPAC 292
DB 235 GYELGGVYAHLPENHYGRVILDAVNPBGADTMGAENQARGFORALDIYESTGQEP--- 291
QY 293 LGTDSAQWVNRNHALVDPLVQKPGKTSDP-RGLSYADATTTGTINALYSPOKMKYLTSGLL 351
DB 292 ----RQGSRKIAGLIERLDAEPLRTSSPQRELQTLAFTGIVLPYLSGMPALISALK 346
QY 352 GIORGSDAGDLVLADYDGRADGHIYSNDQAFNAVKCY--YAETPADPAAMVADQOR 408
DB 347 AAEEG-DGSELLALADGNERDPSGRYGTTHSQRVISCLDCKORPTVETKKLL---PR 402
QY 409 IROVAFELSYGOFTGSAPRDLCAIMVPVATSTPHRAAGAGVUVVAVSTTHDPATYQSG 468
DB 403 FEKVSIV--FGAFLGMDTACGCHDMFVYAGONETAEVSAAPDAFVLYVAGTGDPAITYEGA 460
QY 469 VDLARQLG---APLITFDGTOHTAVFDGNQCVDSAVMHHYFLDGTLPRTSLRCA 518
DB 461 RRAADELGDVGVVLTMOEGHGAYGNGSCVDSAVDAYILKGTVPKDKGVCS 513

RESULT 7
US-08-173-508-2
; Sequence 2, Application US/08173508
; Patent No. 5616485

GENERAL INFORMATION:
APPLICANT: Bartfield, Daniel
APPLICANT: Butler, Michael J.
APPLICANT: Hadary, Dany
APPLICANT: Jenish, David
APPLICANT: Krieger, Timothy
TITLE OF INVENTION: STREPTOMYCES PROTEASES AND IMPROVED
TITLE OF INVENTION: STREPTOMYCES STRAINS FOR EXPRESSION OF PEPTIDES AND
TITLE OF INVENTION: POLYPEPTIDES
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W.
CITY: Washington, D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/173,508
FILING DATE: 23-DEC-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Bent, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 18740/125/CACO
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202 672 5300
TELEFAX: 202 672 5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 537 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-173-508-2

Query Match 20.8%; Score 570.5; DB 1; Length 537;
Best Local Similarity 31.1%; Pred. No. 6.9e-45;
Matches 173; Conservative 82; Mismatches 233; Indels 69; Gaps 21;

QY 5 RRRPSSALLSFGLLGGLLAAPLAGATEEGAGQTP-----GAPVVA 49
DB 2 RKSIRRRATAG--TAGALVATATLAGAVSAPMAAPADGHHGRSMDREARCAIATA 59
QY 50 POOS-----WNSCRETIADTSEIRTAACATVSVVDVDOPGGTQAKLAVIRPATG--O 101
DB 60 ARAARAGIMEDC---AADMNLPKPIQCGYVTVPMDYAKPYKQIRLAVDRIGMTGRSE 116
QY 102 RGCALVNPGRGASAVDVAAAPAIAD-TDILRHFDLVGFDPKGVSHSTPRLKRTDA 160
DB 117 RGALVYNPGRGSGLRPAPRTYTKSAVAMANTAKAYDVGFDPRGVGAP-ISCVDPO 175
QY 161 EF-DAYRRDPMADYSPAGVYTHQVYROLAODCVDRMGFSFLANIGTASVARDMDVNRQA 219
DB 176 EFKAPKAPVPG-SEADKRAQRKILAREVAGCGFEESG-EMLRHMTPTATADLDVIRRA 233
QY 220 LGDDQDINTYGYSTYKLGTAUYLEREGTHYRAVWLDAIDPAVSPITEESISOMAGPOTA 277
DB 234 LGCKKLTNYGVSYGTVLGAIVYGTLPDHRVRYVDSVNPBSRDKIYVQANLDQDVAFEGR 293
QY 278 FNDYAA-DCARSPACPLGTDSDAQWVNRNHALVDPLVQKPGKTSDP------SYAD 328
DB 294 WKDWDVAAADAAYHLGDTRAEVDOVKLRAAAKKRPLGGVYGAELISFFQSAPIYVD 353
QY 329 ATTGTINALYSPOKMKYLTSGLLGIORGSDAGDLVLADY--DGDADAGHYS--NDQDA 384
DB 354 SAMAPTAELFS-----KYY-----AGDTQALVYAAAPDLSDTGMNSAENGNAV 397

OY 385 FNAVBCVAPPTDPAWAAADRIQVAPELSYGOTSGAPSDLCALMPVATSPHPA 444
| : | | : | : | : | : | : | : | : | : | : | : | :
DB 398 YFAVECTDAKWRANMRTWDNRNTRLRDRHPRFMWANAAMNLP---CATPYVKOOTPLINVK 454

OY 445 APAGAKVVVYSTTHDPATRPYOSGVDLARQL-GAPLIT-FDGTOHRVAFPDGNOCVSVM 502
| : : | : | : | : | : | : | : | : | : | : | : | :
DB 455 TSKGLPVLIVQSERDAPTEGAVALHQRFRGRSLITERDAGSHGVTGLVNPCINDRID 514

OY 503 HFFLDGTLPPTSRCAP 519
| : | : | : | : | : | : | : | : | : | : | : | : | :
DB 515 TYLTGRTDARDVCAP 531

RESULT
8
US-08-265-310-2
Sequence 2, Application US/08265310
Patent No. 5856166
GENERAL INFORMATION:
APPLICANT: Bartfeld, Daniel
APPLICANT: Butler, Michael J.
APPLICANT: Hedary, Dany
APPLICANT: Jenish, David
APPLICANT: Krieger, Timothy
APPLICANT: Malek, Lawrence T.
APPLICANT: Soostmeyer, Gisela
APPLICANT: Walczyk, Eva
APPLICANT: Krysgman, Phyllis
TITLE OF INVENTION: STREPTOMYCES PROTEASES AND IMPROVED
TITLE OF INVENTION: STREPTOMYCES STRAINS FOR EXPRESSION OF PEPTIDES AND
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W.
CITY: Washington, D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/265,310
FILING DATE: 24-JUN-1994
CLASSIFICATION: A35
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/173,508
FILING DATE: 23-DEC-1993
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 18740/133/CACO
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202 672 5300
TELEFAX: 202 672 5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 537 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-265-310-2

Query Match	20.8%;	Score 570.5;	DB 2;	Length 537;
Best Local Similarity	31.1%;	Pred. No. 6.9e-45;		
Matches 173;	Conservative 82;	Mismatches 233;	Indels 69;	Gaps 21;
5	RRRPLSSALLISGCGLLAAPPVLAGTPEPGAGQTP-----GAPVYA	49		

[illegible]

US-08-951-742-2
 9
 : Sequence 2, Application US/08951742
 : Patent No. 6127144
 : GENERAL INFORMATION:
 : APPLICANT: Bartfeld, Daniel
 : APPLICANT: Michael J. Butler
 : APPLICANT: Danny Hadary
 : APPLICANT: David Jenish
 : APPLICANT: Tim Krieger
 : APPLICANT: Lawrence T. Malek
 : APPLICANT: Gisela Soostmeyer
 : APPLICANT: Eva Walczyk
 : APPLICANT: Phyllis Krygsman
 : APPLICANT: Shella Garven
 : TITLE OF INVENTION: METHOD FOR EXPRESSION OF PROTEIN
 : TITLE OF INVENTION: BACTERIAL HOST CELLS
 : NUMBER OF SEQUENCES: 25
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Foley & Lardner
 : STREET: 3000 K Street, N.W., Suite 500
 : CITY: Washington
 : STATE: D.C.
 : COUNTRY: USA
 : ZIP: 20007-5109
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: PatentIn Release #1.0, Version #1.30
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/951,742
 : FILING DATE: 16-Oct-1997

ATTORNEY/AGENT INFORMATION:

NAME: Bent, Stephen A.
 REGISTRATION NUMBER: 29,768
 REFERENCE/DOCKET NUMBER: 019740/0140
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (202)672-5300
 TELEFAX: (202)672-5399
 TELETYPE: 904136
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 537 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-951-742-2

Query Match 20.8%; Score 570.5; DB 3; Length 537;
 Best Local Similarity 31.1%; Pred. No. 6,9e-45;

Matches 173; Conservative 82; Mismatches 233; Indels 69; Gaps 21;

5 RRRPSSALLSEGLIGLLAAPLAGATEEPGAGQTP-----GAPVVA 49
 2 RKSSIRRRATAFG--TAGALVTATLLAGANVASAPASAPADGHGHSMDREARGAIAA 59
 50 PQOS-----WNSCRFIADTSEIRTCARCATVSPVVDYDQPGTQALAVIRVATG---Q 101
 60 ARARAGIMEDC---AAMNLPKPIQCGVYVPMYARFKQIRLAVDRIGTGTRE 116
 102 RFGALLVNPFGPGASAVDMVAAPAIAD-TDILRHFDLVGPDPRGVGSHTPALRCRTDA 160
 117 RGQALLYNQGGSGSLRFPARTNKSAYWANTAKAYDVGFPFGVGHSAF-ISCVDQ 175
 161 EF-DAYRDPMDYSPAGYTHVQVYROLAQDCVDRMGSEFLANIGTASYARDMDVYRQA 219
 176 EFVAKRKPDPVPG-SEADKRAQKRLAREYAECCFERSG-EMLPHTTPTMARDLDVIRAA 233
 220 LGDDQINNYIGSYGTLGTRAYLEREGTHRAMWLDGAIIDPVSPT--ESSISMAQFORA 277
 234 LGSKKRLNGLVSGYGLGAVYGTLPFDHVRMYVDVSVNSRDKIYQANLDDQVAFEGR 293
 278 FNDYAA-DCARSPACPGLGTSAGVNRHYALVDPLVQKP-GKTSDDPRL-----SYAD 328
 294 WKQMDQWVAANDAYHLGTRAEVDQWMLKRAAAKKPLGIVGVGAELISFGQAPYD 353
 329 ATTGTINALYSPQRMKYLTLSGLLGLOGSDAGDLVLADY--DGRDADGHS--NDQDA 384
 354 SAWAPRAELIFS---KYV-----ACDTQALVDAAAPDLSDTAGNSAENGNAV 397
 385 FNVNRGVVAPTAPPAWVAADQRIQVAPELSYGQFTGSAPRDLGALMPVPTSTPHRA 444
 398 YTAVECTDAKMPKMTWDRDNLRLHDFEMWAMAMNLP--CATWVFKOOTPLANK 454
 445 APAGAKVVVVSTTHDPATPYQSGVDLARQL-GAPLIT-FDGTQHTAVFDGNCVDSAVA 502
 455 TKGGLPVLVLYGSEDAAPRYEGAVEHLQRFGRSRLITERDGSHEVTGLVNCINDRDV 514
 503 HTFLDGTLPPTSLKCAP 519
 515 TYLLTGTARDAYTCAP 531

RESULT 10

US-08-804-227C-9
 ; Sequence 9, Application US/08804227C
 ; Patent No. 5876991
 ; GENERAL INFORMATION:
 ; APPLICANT: Dehoff, Bradley S.
 ; APPLICANT: Kuhstoss, Stuart A.
 ; APPLICANT: Rostock, Paul R., Jr.
 ; APPLICANT: Sulton, Kimberly L.
 ; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
 ; NUMBER OF SEQUENCES: 15

CORRESPONDENCE ADDRESS:

ADDRESSEE: THOMAS G. PLANT 1501
 STREET: LILLY CORPORATE CENTER
 CITY: INDIANAPOLIS
 STATE: IN
 COUNTRY: USA
 ZIP: 46285
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: MS-DOS
 SOFTWARE: ASCII(DOS) Text only
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/804,227C
 FILING DATE: February 21, 1997
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Plant, Thomas G.
 REGISTRATION NUMBER: 35,784
 REFERENCE/DOCKET NUMBER: X-8231
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 317-276-2459
 INFORMATION FOR SEQ ID NO: 9:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1996 amino acids
 TYPE: amino acid
 TOPOLOGY: unknown
 MOLECULE TYPE: peptide
 US-08-804-227C-9

Query Match 4.0%; Score 108.5; DB 2; Length 1996;
 Best Local Similarity 23.0%; Pred. No. 0.61;

Matches 134; Conservative 50; Mismatches 186; Indels 213; Gaps 32;

26 AAPPLGATEEPGA-----GOTPG---APVAPQOSWNSCRFIADTSEIRTCAR 71
 1426 AAGPEGGADIPGAAWGLVRAQAQOPGRFTLVDDGTQASIRALPGLLATDAGOSAVR 1485
 72 CATVSP--VDYDQ---GTOQAKLAVIRP-ATGQREGALLVNPFGPGASAVDMVAAM 124
 1486 DGRVTVPRVLVPADVPVHGSGTAADGTGAGEPSATLDPREGVLIT-GGIGA---LAAPT 1540
 125 ARAIADTDLRHFDLVGF---DPRGVGH-----STPALRCRTDAFDVYRR--DPM 170
 1541 ARHLVDNRKVRHLLLVAGRGPRAPGVRLVAVELTESGAEVAVRACDVTDRDLRLDLAL 1600
 171 ADYSP-----AGV-----TFVEQYVROLADQCVDRMGF 198
 1601 PDEHPLTCVNHAGVLDGGLVLSQTAERIDTVLRPKADAAYHLDELTRFTRGR--VPLVY 1658
 199 SFL-ANIGTASVA-----RDMDV---RQALGDDQINLYGYSYGTKLQ-----TA 239
 1659 SSVSATLGSAGQGYAANAAMFMDALARRCAAGHPALS-LGWGWSGVGLATGLDQADAA 1717
 240 YLEREG-----THVRAMVLDGAI---DRAVSPRIEESISQMAQFORAFNDYAADCARSPAC 291
 1718 RVRRSGLAPLDAGALDLDRALTRPEPALLPVRLDLRAAG-ATALEPVLDLGVPA- 1775
 292 PLGTDSAGWNRHYHALVDPLVQKPKTSDPRGLSYVADATGTINLSPQRMKYLTLSGL 351
 1776 -----DARSTPGAAAGTG----- 1788
 352 GLQSGDAGDLVLADYDGRDADGHSNDODAFNVRVCVAPTPADPAWVAADQRIQ 411
 1789 -----DEDG-----AVR--PAPAPADAAGTLAA----- 1809
 412 VAPFLSYGQFTG-SAPRDLCAIMPVPATSTPHRAPAGKXVVVVS---TTHDPATPYQS 467
 1810 -----FLAGRSAPERTALLDLVITEV---AAVLGHGDPAALGAARTFMDAGFDSL 1858
 468 GVDLARQLGAPLITFDGTO--HTAVFDGNCVDSAVMNHFFLNG 508

Db 1859 AVDLNRNLN-----TRTGLRLPATLVFD--HPTPLALAEILLDG 1895

RESULT 11

US-08-804-198-3
; Sequence 3, Application US/08804198
; Patent No. 5945320

GENERAL INFORMATION:

APPLICANT: Burgett, Stanley G.
APPLICANT: Kunsloss, Stuart A.
APPLICANT: Rao, Nagaraja R.
APPLICANT: Richardson, Mark A.
APPLICANT: Rostek, Paul R., Jr.
TITLE OF INVENTION: PLATENOLIDE SYNTHASE GENE
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: PAUL R. CANTRELL 1138
STREET: LILLY CORPORATE CENTER
CITY: INDIANAPOLIS
STATE: IN
COUNTRY: USA
ZIP: 46285

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: Macintosh
OPERATING SYSTEM: Macintosh 7.0
SOFTWARE: Microsoft Word 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/804,198
FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
NAME: CANTRELL, PAUL R.
REGISTRATION NUMBER: 36,470
REFERENCE/DOCKET NUMBER: P9113
TELECOMMUNICATION INFORMATION:
TELEPHONE: 317-276-3885
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1996 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-804-198-3

Query Match 4.0%; Score 108.5; DB 2; Length 1996;

Best Local Similarity 23.0%; Pred. No. 0.61; Mismatches 186; Indels 213; Gaps 32;
Matches 134; Conservative 50;

QY 26 AAPPLAGATEEPGA-----GQTPG-----APVVAPOOSWNSCREFIADTSEIRTA 71
Db 1426 AAGEGGADPAGAAVGLRVNAEQRFRTLVVDGTQASLRALGGLATDAGOSAVR 1485
QY 72 CATYVSV--VDYDOP-----GTOAKLAVIRYP-ATGQRFGLLVNPGPGASAVDMAAM 124
Db 1486 DGRATVRLPVADPVHGGTADGTAGEPFSATILDPGTVILT--GGTGA---LAAET 1540
QY 125 APAIADDIIRHFDLVG---DPRGVGH-----STRALCRDPAEDPAVR--DPM 170
Db 1541 ARHLVDHAKVRHLLLVGRGPDAGVRLVAELTESGAEVAARACDVTDRDLRRLLDAL 1600
QY 171 ADYSP-----AGV-----THVBOYROLAODCVDRMGF 198
Db 1601 PDEHPLCVVHTAGVLDVGLSAQTAERIDTVLRPKADAHVHDELRLRIGR--VPLVLY 1658
QY 199 SFL-ANIGTASVA-----RDMNV---RQALGDDQINITYGISTGTLG-----TA 239
Db 1659 SSVSATLGSAGAGYAANAFMDALARRCAAGHPALS-LGWMGWSGVGLATGLDGA 1717
QY 240 YLEBFG-----THVRAVNLGAI---DPAVSPITESSISQMAFOTANNDVAADCARSPAC 291
Db 1718 RVRSSGLAPLDAGAAIDLDRALTRREPALLPVRLDRAAG--ATALPEVLRDLAVPA- 1775

QY 292 PLGTDSAQWNRHYHALVDPVQKPKTSDPRGLSYADATTGTTINALYSPORWKYLTSG 351
Db 1776 -----DARSTPGAAAGTG----- 1788
QY 352 GLORGSAGDLVLVADDDYDGRDADGHSNDODAFNAVRCYAPPTPAPAAVADOKIRQ 411
Db 1789 -----DEDG-----AVR--PAPAPADAAGTLA----- 1809
QY 412 VAPFLSYGQFTG--SAPRDLCALMPVPATSPHPAPAGAGKVUVS---TTIHDPATPYQS 467
Db 1810 -----RLAGRSAPERTALLDLVRETV---AAVLGCHDPAATGAAKRTFKDAGFDSL 1858
QY 468 GVDLARQLAGPLITFDGTQ--HTAVFDGNOCVDSAVMHYFLDG 508
Db 1859 AVDLNRNLN-----TRTGLRLPATLVFD--HPTPLALAEILLDG 1895

RESULT 12

US-09-320-878-1
; Sequence 1, Application US/09320878A
; Patent No. 6117659

GENERAL INFORMATION:

APPLICANT: ASHLEY, Gary
APPLICANT: BETLACH, Melanie C.
APPLICANT: BETLACH, Mary C.
APPLICANT: MCDANIEL, Robert
APPLICANT: TANG, Li
TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
FILE REFERENCE: 300622002120
CURRENT APPLICATION NUMBER: US/09/320,878A
CURRENT FILING DATE: 1999-05-27
EARLIER APPLICATION NUMBER: CIP OF 09/141,908
EARLIER FILING DATE: 1998-08-28
EARLIER APPLICATION NUMBER: CIP OF 09/073,558
EARLIER FILING DATE: 1998-05-06
EARLIER APPLICATION NUMBER: CIP OF 08/846,247
EARLIER FILING DATE: 1997-04-30
EARLIER APPLICATION NUMBER: 60/119,139
EARLIER FILING DATE: 1999-02-08
EARLIER APPLICATION NUMBER: 60/100,880
EARLIER FILING DATE: 1998-09-22
EARLIER APPLICATION NUMBER: 60/087,080
EARLIER FILING DATE: 1998-05-28
NUMBER OF SEQ ID NOS: 34
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 4551
TYPE: PRT

US-09-320-878-1
; ORGANISM: Streptomyces venezuelae

Query Match 3.9%; Score 106.5; DB 3; Length 4551;

Best Local Similarity 21.1%; Pred. No. 3.4; Mismatches 223; Indels 185; Gaps 27;
Matches 125; Conservative 60;

QY 8 PLSSALTSF-----GLLT-GGILLAAPLAGATEEPGAGQTPGAAPVVAPOOSWNSCREFI 62
Db 3518 PLGAVVSLPDRDGLLTLTGRLSLRTHPWLADHAVLGSVLLPGAAV-----ELA 3570
QY 63 DTSEIRTARCATVSVPVYDOPGTOAKLAVIRVPATGQRFGLLVNPG---GPGASAV 118
Db 3571 ESAGLRQVRELTLEPRLVLRHGVGVELRVTV-----GAPAGPGEASAGDGARV 3620
QY 119 DMAAAMPALADIDILRHFDLVGFDPRGV---HSTPALCRDPAEDDAYRRPMDYSP 175
Db 3621 SLHSRLADA-----PAGTAMSCHATGTL--ATDRBELPVAPDRAAMMP 3662
QY 176 AGVTHV--EQVYROLAQDCVDRMGFSFLANIGTASVARDMDVROALGD-----DQ 225
Db 3663 QGAEVPLDGLYERL-----DGNGLAF-----GPLFQGINAVRYEGEVFADIALPAT 3711

```

ANTI-SENSE: NO
FRAGMENT TYPE:
ORIGINAL SOURCE:
ORGANISM: Streptomyces coelicolor
STRAIN: N/A
INDIVIDUAL ISOLATE:
DEVELOPMENTAL STAGE:
HAPLOTYPE: N/A
TISSUE TYPE: N/A
CELL TYPE: unicellular organism
CELL LINE:
ORGANELLE:
IMMEDIATE SOURCE: N/A
POSITION IN GENOME:
FEATURE:
NAME/KEY: AbsA1
LOCATION:
IDENTIFICATION METHOD: deduced
OTHER INFORMATION:
PUBLICATION INFORMATION:
AUTHORS:
TITLE:
JOURNAL:
VOLUME:
ISSUE:
PAGES:
DATE:
RELEVANT RESIDUES IN SEQ ID NO:
US-08-796-414B-1

Query Match 3.9%; Score 106; DB 2; Length 571;
Best Local Similarity 23.7%; Pred. No. 0.15;
Matches 93; Conservative 38; Mismatches 117; Indels 144; Gaps 22.

QY 176 AGVTHVEQV-YKQ-----LAQDCVDMRG--FSPLANI-GTASVARDM-DMV 216
Db 170 AGWERAQLEWRKORYIADQARKERARIAODIHDSIGHELVSMAILLAGLELAPGLSDPH 229
QY 217 ROALGDDDDINVLGVSYGTKLGTAVERFGTHVRAMVLDGALIDPAVSPTEESISOM----- 271
Db 230 RESVG--QLR-----ERCTMATERL--HEVIGLKRDPNPSTLPADSVYAOLVRRFQ 277
QY 272 -AGFQTAFNDYAADCARs-PACPLGTDSASQWVNRHVALVDPLVQKPGKTSDPRLGSLVADA 329
Db 278 RSGFVRRQE--DQARBRPGTPLLSDLAAYRVVQGEALITNMAKHMPGAPIDVRVHTHDAE 334
QY 330 TT-----GTINALYSPQRKKYTLTSLGLG-----LQKSGDAGDLLVLA- 366
Db 335 TVVSVYNERPERGGSVPAAGS-----GSGILIGLDERVRLAGTGLRTPRAGCFEYVAR 387
QY 367 -----DDYDGRDADGHYSNDQAFNVRQVYA-----PTPADP----- 399
Db 388 LPRGASSPSRSSTEPKPADGCGTAGGSGDTAPGAATAGNENGRAAAAADLPAPSGPWRS 447
QY 400 ---AAMVADQRIRO-----VAPFLSYCOF-----TGSAPRDLCAL--- 432
Db 448 ASRAALLLTRIRIRRDARRALLIPAVLGAIALYATFG-GLIYFTSATITSLAPEDYARIKVG 506
QY 433 -----WEPVATSTPHAPAPAG 448
Db 507 ETRADLAPALPERIRIKKPPVTS--EPSVPAG 536

RESULT 14
US-09-335-409-10
; Sequence 10, Application US/09335409
; Patent No. 6121029
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 5, 2001, 14:26:28 ; Search time 62.98 seconds
(without alignments)
281.781 Million cell updates/sec

Title: US-09-461-774-4

Perfect score: 2738

Sequence: 1 AAMRRRRRLSSALLSFGILL.....AVMHYFLDGLTPPTSIRCAP 519

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 268485 seqs, 34193795 residues

Optimal number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq.36:*
1: /SIDSL/gcgcdata/geneseq/geneseqp/AA1980.DAT:*
2: /SIDSL/gcgcdata/geneseq/geneseqp/AA1981.DAT:*
3: /SIDSL/gcgcdata/geneseq/geneseqp/AA1982.DAT:*
4: /SIDSL/gcgcdata/geneseq/geneseqp/AA1983.DAT:*
5: /SIDSL/gcgcdata/geneseq/geneseqp/AA1984.DAT:*
6: /SIDSL/gcgcdata/geneseq/geneseqp/AA1985.DAT:*
7: /SIDSL/gcgcdata/geneseq/geneseqp/AA1986.DAT:*
8: /SIDSL/gcgcdata/geneseq/geneseqp/AA1987.DAT:*
9: /SIDSL/gcgcdata/geneseq/geneseqp/AA1988.DAT:*
10: /SIDSL/gcgcdata/geneseq/geneseqp/AA1989.DAT:*
11: /SIDSL/gcgcdata/geneseq/geneseqp/AA1990.DAT:*
12: /SIDSL/gcgcdata/geneseq/geneseqp/AA1991.DAT:*
13: /SIDSL/gcgcdata/geneseq/geneseqp/AA1992.DAT:*
14: /SIDSL/gcgcdata/geneseq/geneseqp/AA1993.DAT:*
15: /SIDSL/gcgcdata/geneseq/geneseqp/AA1994.DAT:*
16: /SIDSL/gcgcdata/geneseq/geneseqp/AA1995.DAT:*
17: /SIDSL/gcgcdata/geneseq/geneseqp/AA1996.DAT:*
18: /SIDSL/gcgcdata/geneseq/geneseqp/AA1997.DAT:*
19: /SIDSL/gcgcdata/geneseq/geneseqp/AA1998.DAT:*
20: /SIDSL/gcgcdata/geneseq/geneseqp/AA1999.DAT:*
21: /SIDSL/gcgcdata/geneseq/geneseqp/AA2000.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	780	28.5	539	16	R80506
2	780	28.5	539	20	W87799
3	653.5	23.9	513	20	W87798
4	647.5	23.6	513	16	R80504
5	570.5	20.8	537	16	R80508
6	570.5	20.8	537	16	W87796
7	128	4.7	785	19	W72103
8	128	4.7	3119	19	W72204
9	123	4.5	1051	19	W52304
10	116	4.2	1291	19	W52912
11	113.5	4.1	459	12	R13498
12	111	4.1	732	19	W72061

13	108.5	4.0	1996	18	W22607
14	108.5	4.0	1996	18	W23717
15	108	3.9	1638	20	Y00138
16	108	3.9	1638	20	Y00140
17	108	3.9	1638	20	Y00142
18	108	3.9	12199	21	Y77180
19	106.5	3.9	4551	21	Y67201
20	106.5	3.9	4613	21	Y77192
21	106.5	3.9	4613	21	Y77200
22	106	3.9	571	20	W78491
23	104	3.8	402	21	Y77276
24	104	3.8	402	21	Y78828
25	104	3.8	423	21	Y58581
26	102.5	3.7	574	18	W23300
27	102.5	3.7	613	20	Y05278
28	102	3.7	534	20	W93007
29	102	3.7	920	21	Y77291
30	102	3.7	920	21	Y78843
31	101.5	3.7	590	18	W31751
32	101.5	3.7	590	19	W71888
33	101	3.7	582	15	R42306
34	101	3.7	928	21	Y77292
35	101	3.7	928	21	Y78844
36	101	3.7	3724	18	W23718
37	101	3.7	3724	18	W22608
38	100.5	3.7	1978	20	Y27230
39	99.5	3.6	1021	19	W76192
40	99	3.6	311	18	W22843
41	98	3.6	913	19	W53829
42	98	3.6	850	19	W62040
43	97.5	3.5	4630	18	W19629
44	97	3.5	4630	18	W19629
45	96.5	3.5	2115	19	W59276

ALIGNMENTS

RESULT 1	
ID R80506	R80506 standard; Protein: 539 AA.
XX	
AC R80506;	
XX	
DT 04-DEC-1995 (first entry)	
XX	
DE S. lividans protease P5-6.	
XX	
KW Protease: metalloendoproteinase; tripeptidyl aminopeptidase;	
KM protease-deficiency; protein secretion.	
XX	
OS Streptomyces lividans.	
XX	
FH Key	Location/Qualifiers
FT Peptide	1..47 /label= Sig_peptide
XX	
PN W09517512-A.	
XX	
PD 29-JUN-1995.	
XX	
PF 22-DEC-1994; 94MO-US14772.	
XX	
PR 23-DEC-1993; 93US-0173508.	
XX	
PA (CANG-) CANGENE CORP.	
XX	
PI Bartfeld D, Butler MJ, Hadary D, Jenish DL, Krieger TJ;	
PI Malek LF, Soostmeyer G, Walczyk E;	
XX	
DR WPI: 1995-240673/31.	
DR N-PSDB: Q99366.	
XX	

Platenolide synth
Platenolide synth
Enterococcus faeca
Enterococcus faeca
Enterococcus faeca
S. venezuelae pik
Narbonolide synth
S. venezuelae macr
S. venezuelae pik
S. coelicolor AbsA1
Streptomyces cina
Amino acid sequenc
Sorangium cellulos
Scopulariopsis bre
S. brevicaulis bet
Z. mays ckl1 prote
Streptomyces cina
Ksq-Ang loading di
H chain subunit of
Anti-human Ras mon
Phospholipase D-P.
Streptomyces antib
Ksq-Ang loading di
Platenolide synth
Platenolide synth
Amino acid sequenc
Actinoplanes sp. a
Xanthomonas campes
Pseudomonas Orfx s
Pseudomonas alcali
protein isolated f
Streptomyces venez
Rubella virus RA27


```

Db 362 egawgqlresltsaikedgaqllllsdsyereadgyynlmaaaavcltd-----p 416
QY 400 AANVAADQRIROYAPFLS-----YGFETGSAPRDL-CALMPVPATSTPHRAPAGAGKV 453
Db 417 aatsspe-vrdalpdfekaspvfe--glawsslncaaypvkprtgphrleaaatpiv 473
QY 454 VVESTPDPAFPGSGVDLAQL-GAPLITFDGQHTAVFGNOCVDSAVMHYFLDGLTP 512
Db 474 vvglttdpatpryaaalsqlltsghlllyegdghtaayrgssclsdaIntlyllgtape 533
QY 513 TSLRCA 518
Db 534 dgkrcc 539

RESULT 3
W87798
ID W87798 standard; Protein; 513 AA.
W87798;
19-MAR-1999 (first entry)
Aminopeptidase SIpe encoded by clone p8-2.
DE Aminopeptidase SIpe encoded by clone p8-2.
XX
XX Tripeptidyl aminopeptidase; TAP; N-terminal cleavage; protein production;
KM GM-CSF; Interleukin-3; IL-3; IL-6; EPO; tumour necrosis factor; TNF;
KW SCF; IL-7; IL-2; p8-2; SIpe.
XX
XX Streptomyces lividans.
OS US5856166-A.
XX
XX 05-JAN-1999.
PD 24-JUN-1994; 94US-0265310.
XX
XX 24-JUN-1994; 94US-0265310.
PR 23-DEC-1993; 93US-0173508.
XX
XX (CANG-) CANGENE CORP.
PA
XX Bartfeld D, Butler MJ, Garven S, Hadary D, Jenish DL;
PI Krieger TJ, Krygsmann P, Malek LT, Soostmeyer G, Walczyk E;
XX WPI; 1999-105117/09.
DR N-PSDB; V84067.
XX
XX Streptomyces tripeptidyl aminopeptidase - useful for removing
PT N-terminal pro-peptide from secreted proteins
XX
XX Claim 6; Fig 29A-C; 83pp; English.
XX
XX The present sequence represents a tripeptidyl aminopeptidase designated SIpe. The
CC specification also describes a tripeptidyl aminopeptidase (TAP) of
CC Streptomyces. The aminopeptidase is endogenous to Streptomyces and
CC cleaves an N-terminal sequence of X-Pro-Y, where X is an aliphatic
CC or hydroxy amino acid and Y is an aliphatic, hydroxy or
CC sulphur-containing amino acid. The TAP of Streptomyces are useful in
CC the production of proteins, such as GM-CSF, interleukin-3 (IL-3),
CC IL-6, EPO, tumour necrosis factor (TNF) SCF, IL-7 and IL-2.
XX
XX Sequence 513 AA;
SQ

```

Query Match 23.9%; Score 653.5; DB 20; Length 513;
 Best Local Similarity 33.6%; Pred. No. 2,2e-49;
 Matches 179; Conservative 69; Mismatches 224; Indels 61; Gaps 18;

QY 18 LLLGLLLAAPPAGATEE-----PGAGTGPAPVPAPOQS--WNSCREFIADT 64
 Db 10 ltaaaalltagcgssdedkdkggrssagpsaaapsyaaalsqtlldwarc-----eg 64

```

QY 65 SEIRTA-----RCATVSPVDTYDQPGSTQAKLAVIRVPATG-QRFGALLVNPGRPAASAV 118
Db 65 sddaparpdgqwtcatlkapi:dw:dpdgetldlitsraagddrlsgllfnfngpaaasy 124
QY 119 DWAAAPALADT--DLRHFDLVGPPRGVGHSTALRCRTAPAEFADYRRDMADQSPA 176
Db 125 stm----psydavssiherydlvswdprgva-asegvrtdeaaesvstlpspa 179
QY 177 GVTNVEQVYROLAD----CVDRMGFSFLANIGTASVARDMDVNRALGDQJINTGYLSY 232
Db 180 -----eeqaylkdaadfggrgekaag--klmehvsttdardmdlmrlnvldermhyfyisy 234
QY 233 GTRLGTAFLERFGSTHVRAMVLDGALDPVAPSPLESISQMGAFOTAPNDYAADCASRPAC 292
Db 235 gteiggyvahlfpehvgrylldavvpgadtmghaenqargfgraldyiescgper--- 291
QY 293 LGTDSAQWVNRHAYLDPLQKPGKTSDP--RGLSYADATGTGTINALXSPQRMKYTLTSGLL 351
Db 292 -----egsrklagllleridaerlpsspgreltqlatlgivlpilysesgwpallsaik 346
QY 352 GIORGSDAGDLVLADYDGRDADGHYSNDQAFNNAVCV---YAPTPADPAAMVAADQR 408
Db 347 aaeeg-dgsellaladgynerdpsgrgtthsgrylsclddkqrpvtcekkll---pr 402
QY 409 IROYAPPLSTGFTGSAPRDLALMPVPATSTPHRAPAGAGKVUVVSTTHDPATPYQS 468
Db 403 fekvspv--fgatlgwdtagwchndwprvqghetaesapdaarvllvngtgdpatryge 460
QY 469 VDLAROLG---APLITFDGQHTAVFDGNOCVDSAVMHYFLDGLTPPTSLRCA 518
Db 461 rrmadelgkvagvvlwlcgeghnagysocvsadavdllykvtvpkdkvcs 513

```

```

RESULT 4
R80508
ID R80508 standard; Protein; 513 AA.
AC R80508;
XX
XX 04-DEC-1995 (first entry)
DT
XX
XX S. lividans protease p8-2.
DE
XX
XX Protease; metalloendoproteinase; tripeptidyl aminopeptidase;
KM protease-deficiency; protein secretion.
XX
XX Streptomyces lividans.
OS
XX
XX WO9517512-A.
PN
XX
XX 29-JUN-1995.
PD
XX
XX 22-DEC-1994; 94WO-US14772.
PF
XX
XX 23-DEC-1993; 93US-0173508.
PR
XX
XX (CANG-) CANGENE CORP.
PA
XX Bartfeld D, Butler MJ, Hadary D, Jenish DL, Krieger TJ;
PI Malek LT, Soostmeyer G, Walczyk E;
XX WPI; 1995-240673/31.
DR N-PSDB; Q993368.
XX
XX Endogenous Streptomyces protease(s), opt. having impaired activity -
PT useful in prodn. of exogenous proteins with reduced proteolytic
PT degradation
XX
XX Claim 4; Fig.19; 142pp; English.
PS
XX
XX Protease TAP-negative cells were transformed with a S. lividans 66
CC genomic library and screened with APA-beta-naphthylamide to

```


QY	503	HYFLDGLTPEPTSLRCA	519		
DB	515	tylltgrtdardvtcap	531		
RESULT 6					
ID	W87796	W87796 standard; Protein; 537 AA.			
AC	W87796;				
DT	19-MAR-1999	(first entry)			
XX					
DE	Tripeptidyl aminopeptidase (tap) protein.				
XX					
KW	Tripeptidyl aminopeptidase; TAP; N-terminal cleavage; protein production;				
KW	GM-CSF; Interleukin-3; IL-3; IL-6; EPO; tumour necrosis factor; TNF;				
KW	SCF; IL-7; IL-2.				
XX					
XX	Streptomyces lividans.				
FT	Key	Location/Qualifiers			
FT	Misc-difference 1	/note="Met encoded by TTG"			
FT	Peptide	1..39			
FT		/note="signal peptide"			
FT	Protein	40..537			
FT		/note="mature protein"			
XX					
PN	US5856166-A.				
XX					
PD	05-JAN-1999.				
XX					
XX	24-JUN-1994;	94US-0265310.			
PF					
XX	24-JUN-1994;	94US-0265310.			
XX					
PR	24-JUN-1994;	94US-0265310.			
PR	23-DEC-1993;	93US-0173508.			
XX					
PA	(CANG-) CANGENE CORP.				
PI	Bartfeld D, Butler MJ, Garven S, Hadary D, Jenish DL;				
PI	Krieger TJ, Kryszman P, Malek LP, Soostmeyer G, Malczyk E;				
XX					
XX	WPI; 1999-105117/09.				
DR	N-PSDB; V84065.				
XX					
XX	Streptomyces tripeptidyl aminopeptidase - useful for removing				
XX	N-terminal pro-peptide from secreted proteins				
PS	Claim 2; Fig 12A-B; 83pp; English.				
XX					
CC	The present sequence represents a tripeptidyl aminopeptidase (TAP) of				
CC	Streptomyces. The aminopeptidase is endogenous to Streptomyces and				
CC	cleaves an N-terminal sequence of X-Pro-Y, where X is an aliphatic				
CC	or hydroxy amino acid and Y is an aliphatic, hydroxy or				
CC	sulphur-containing amino acid. The TAP of Streptomyces are useful in				
CC	the production of proteins, such as GM-CSF, interleukin-3 (IL-3),				
CC	IL-6, EPO, tumour necrosis factor (TNF) SCF, IL-7 and IL-2.				
XX					
SQ	Sequence 537 AA;				
Query Match 20.8%; Score 570.5; DB 20; Length 537;					
Best Local Similarity 31.1%; Pred.No. 4.7e-42;					
Matches 173; Conservative 82; Mismatches 233; Indels 69; Gaps 21					
QY	5	RRRPSALLSLGGLLAPLAGATEEPGAGQTP-----GAPVVA	49		
DB	2	rkssirratatg--tagtlvtatliagvsapaasaadpghygrwsdreargaiaa	59		
QY	50	PGQSS--WNSCREFIADTSEIRARCAVSVVDYDQPGGTQAKLAVIRVPATG--Q	101		

D6	aaaraagatidwecd---aadnmlrprikqgytvtmmdydkrygkaiirlavdrldgnltgluse
OY	102 REGALLVPGGPRGASAVDMVAAMAPALAD-TDILRHFDLYGFEDRQGHSTPALRCRTDA
D6	117 rgallvpggprrgsqglrlfrparvtlnksawantakaydfvgfdprgvghsep-lscvdpq
OY	161 EF-DLVRRPDPAADVSPAGVTNHVEGYROLADDCVDRLMGCSFLANIGTAASVAROMDWROA
D6	176 efvkarkpadrvpyr-seackraqrkrlareyaegcfezsg-emlpmntcpnrtdivlrtae
OY	220 LGDDOINLGVSYGTGLKLTATYLEEFGTTHRAMVLDTGAIDPRAVSPI--EEISQMAGFOTA
D6	224 lgekltnlydvrsygtlygaavgclfrdhvrtmtvvdsvsnpesrdkiwygdanldqdvaltegr
OY	278 FNDYAA-DCANSPPACPLCTDSQAQWVNRYHALVDPLVGKP-GKTSDPRGL-----SVAD
D6	294 wkdwqvwvaandaahylhgdttraeydqqlklraaackplrgyvvpaelisffgsapyud
OY	329 ATTGITNMLSPQRKKYLITSLGLDLGRSDBAGDLLVLAADDY--DGRRADGNHS--NQDA
D6	334 sawaprlaeifs---kyv-----aqdcqaivaldaaprslstagnasaengnav
OY	385 FNAYGCVPVAPPDPADPAAWAADORLRQVAFPLFSYGQFTSGARPOLCALMPVATSHPHA
D6	398 ylavectckakpraanvrtrdntrtlrhphfimtvanamnmr---catwpkkqqlpnlvk
OY	445 APAGAKVVVVYSTHDPTPYQSQGVLDLAROL-GAPLLTP-FDGTONTAVTFDNOCVDASVM
D6	455 lgkgjlprrllvgserdaetpyegeavelhqrftrsrllderagshytgvlvnpeindrvd
OY	503 HYFLDGIPTPTSUKCAP
D6	515 tylltgtirdardvtcap
RESULT	7
ID	w72103 standard; Protein; 785 AA.
XX AC	w72103;
XX DT	18-DEC-1998 (first entry)
DE XX	HSV-2 strain SB5 Contig ID 11 ORF#1 protein.
KM KW	HSV-2 strain SB5; Immunological response induction; therapy; antiviral identification; viral protein inhibitor.
OS XX	Herpes simplex virus type 2.
FH FH	Key Location/Qualifiers Misc-difference 780..785 /note= "encoded by C, unspecified amino acids"
FN PD	w09820016-A1. 14-MAY-1998.
PE PE	31-OCT-1997; 97MO-US20016.
PR PR	09-JUN-1997; 97US-0049018. 04-NOV-1996; 96US-0030279.
PA PI	(SMIK) SMITHKLIN BEECHAM CORP.
DR DR	N-P5DB; V62155.
PT PT	Herpes simplex virus type-2 sequences - useful in, e.g. prevention and treatment of infection or inducing immunological response in


```

Db      2643 sppdaalpppatsgsaafts 2663

RESULT 9
ID      W52304 standard; Protein; 1051 AA.
XX
AC      W52304;
XX
DT      23-JUN-1998 (first entry)
XX
DE      Glucodextranase protein.
XX
KW      Glucodextranase; enzyme; dextran.
XX
OS      Arthrobacter globiformis.
XX
FH      Key Location/Qualifiers
FT      Misc-difference 65 /note= "encoded by CTC"
          Misc-difference 86 /note= "encoded by CTC"
          /note= "encoded by CTC"
          JPI0070981-A.
          17-MAR-1998.
          30-AUG-1996; 96JP-0230128.
          30-AUG-1996; 96JP-0230128.
          30-AUG-1996; 96JP-0230128.
          30-AUG-1996; 96JP-0230128.
          (NORO ) NORINSUISANSO SHOKUHN SOCO.
          (NODA ) ZH NODA SANGYO KAGAKU KENKYUSHO.
          WPI: 1998-233629/21.
          N-PSDB: V19895.
          This sequence is the glucodextranase of Arthrobacter globiformis of
          the invention. A transformant containing a vector comprising the DNA
          sequence can be used to prepare glucodextranase efficiently in a short
          period without the requirement for the addition of expensive dextran.
          Sequence 1051 AA:

Query Match      4.5%; Score 123; DB 19; Length 1051;
Best Local Similarity 21.8%; Pred. No. 0.031;
Matches 148; Conservative 62; Mismatches 219; Indels 250; Gaps 37;

```

```

Db      434 ake1kpaady1vaagrktpgetweecgysstslasqiaa1aaagdaekng-----dv9 488
QY      206 TASVARDMDMVRQALGDD-QINYLGYSTGK-----LGTAVL-----BR 243
Db      489 saaiyr-----atadewgrotekwmftfnpgvgdnyllrtisgsgnrvndgaarwmgn 541
QY      244 FGRHVRAMVLDGAI-----DPAVSPTEESISOMAGFQTFNFNDYAADCARSPACPIG 294
Db      542 agvhprenavldggflefvlrvkprpadahvadsiae-----tdaslsqetpg9 589
QY      295 TDSAGQWNNRYHALVDPLVQPKGKTSDPRLSTYDATTGTINALYSPQRMKYLTSGLLGO 354
Db      590 ---rmw-hry-----tydgygekad--gsrwdgtvgvrlwrl]sgergey-----ala 631
QY      355 RGSADGDLVLAD-----DYDGRDADCHYNSDDAFNAVBCVYAPRPADPA 400
Db      632 ngqda1pyla1tmhsaangfmlpegywdqdeptsyghelgrstg-----saspl 680
QY      401 AMVAADQRIQVAPFLSYQFTGS---APRDLCALM-----PVPATSTPHPAAPAGAG 450
Db      681 swama-qyvr-----laaglkngspvclpknvqryasgavgkpal1inspaqltadsa 734
QY      451 KVVVVSST-----HDPATPYQSG-----VDLARQLGAPLITFDGTOHTAWE 491
Db      735 nltvsgttaakkyvsvngnlfap1ltpagdgts1svda-----lpatkntlv9 785
QY      492 ----DGNOCVDS-AVMHY 504
Db      786 aavgsdgtavestrvlhy 804

RESULT 10
ID      W59912 standard; Protein; 1291 AA.
XX
AC      W59912;
XX
DT      20-NOV-1998 (first entry)
XX
DE      Amino acid sequence of the mutanase enzyme.
XX
KW      Mutanase enzyme; PT box; alpha-1,3 glucoside bond; mutan; plaque;
          bacteria; teeth.
XX
OS      Bacillus sp.
XX
FH      Key Location/Qualifiers
FT      Peptide 1..31
FT      Domain 306..354
FT      /note= "PT box"
          JPI0201483-A.
          04-AUG-1998.
          01-OCT-1997; 97JP-0284362.
          25-NOV-1996; 96JP-0314057.
          (LIOY ) LION CORP.
          WPI: 1998-474495/41.
          N-PSDB: V53575.
          This is the amino acid sequence of the mutanase enzyme, comprising
          the a PT box, which decomposes the alpha-1,3 glucoside bond of mutan.
          The mutanase enzyme is used in the method of the invention for

```

CC prevention and removal of plaque and bacteria on teeth
XX
SQ Sequence 1291 AA;

SQ Sequence 1291 AA;

Query Match	4.2%	Score 116	DB 19	Length 1291
Best Local Similarity	18.3%	Pred. No. 0.17		
Matches 97	Conservative 57	Mismatches 195	Indels 180	Gaps 18

```

QY 18 LLLGGLLAARPLAGAAEEPCAGCTCAPVAVPQOSNNSCREFJALDELIRTRACATVSV 77
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 14 mlmsalltavgpfrgpaasaggpnltprkprltasgsgqfypgnvkddgnqlywestinaf 73

QY 78 P--VDYQPGGTAKKAVIRPARGQ--RFGALLVNPFGSPGASAVDMYAA-----MAPIAD 130
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 74 pqwlvqvdligastldgdlvklkipsawearltqlavgslnstftdlygsanayfsvpgn 133

QY 131 TDILRHFDLVGFDPDRGVSHSTPRARCTDA-----EFDAYRDRPMADYSPACVTH 180
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 134 ntlvlnf-----tatslyrtrlvltantltgpaaglaefely---gsqddqtlrppdty 181
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 181 VEQYVROLADQCDVDRMGESFLANTIGTASVARDMDVRLQALDDOINLYGTSYGTGCTAY 240
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 182 qae-----saalsgagkvnltl-----haqy-----lgtgf 206

QY 241 LEREGTHVRAWVLGDALDPAVSPIEESISQMGQGTAFNDYAADCARSPACPLCTDSAQW 300
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 207 vdgywtq-----gattfs----- 220

QY 301 VNRYHALVDPVLRQPKGTSDBRGSLVYADATTG-----TINALYSPORWK 344
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 221 -----vnaptagndvltlrgnabatsnktvslungaklrqfcltprlwnws 269

QY 345 YLTSGLLGLQSGSAGGLLVLADYDGRDADGHSNODAFNAVRCYAAATPRADPAAWVA 404
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 270 skte-clnlagsn-----tlaykydpqds-gnnvldqitveastsrlrppr 314

QY 405 ADORIQVAPFLSYGQETGSAPRDLCAIMVRPAVSTPNHAPRAGACGVVVVSTTHDPATP 464
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 315 -----sprlrprlrppr-clrprlrprlrprlrprlrprlrprlrprlrprlrprlr 353

QY 465 YQSGVDLARQLGAPLITFDGIGNHAYVDGQDCVDSAVMHNFLDGTLPPT 513
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 354 pygnlaigksta-----sshqkyvaeanahdvnlkwegygnpst 395

```

RESULT 11

R13498 standard; Protein; 459 AA

AC R13498;

DT 25-OCT-1991 (first entry)

P.denitrificans COB G.

cor gene; descobaltocorrinoid; cor gene

05 *Pseudomonas denitrificans*.

PN W09111518-A.

PD 08-AUG-1991

PF 30-JAN-1991; 91WO-FR00054.

PR 31-JAN-1990; 90FR-0001137.

PA (RHON) RHONE-POULENC BIOCH.

PI Blanche F, Meron B, Crouzet J, Debussche L, Levy-Schil S;

XX

DR WPI; 1991-252650/34
DR N-PSDB; Q13285.

DR N-PSDB; Q13285.

PT New polypeptide(s) involved in cobalamin and cobamide biosynthesis - and DNA encoding them, for amplification of PT cobalamin, esp. coenzyme B12 prodn.

PS Claim 17; Fig 16; 299pp; French.

This sequence corresponds to one of 24 polypeptides obtained from P. dentrificans and implicated in the biosynthesis of cobalamins and/or cobamids. C086 is specifically involved in the conversion of 5'-deoxy-5'-adenosyl cobyrinic acid a,c-diamide. It is encoded by part of the 8.7kb EcoRI-EcoRI fragment of plasmid pUX57. The plasmid was isolated from a P. dentrificans genomic DNA bank constructed in vector pUX59. See Q13284-Q13288.

CC See Q13284-Q13288

Sequence 459 AA

Query Match	4.1%	Score 113.5	DB 12	Length 459
Best Local Similarity	23.2%	Pred. No. 0.068		
Matches 89	Conservative 49	Mismatches 152	Indels 93	Gaps 20

```

QY      22  GLLAAAPRLGATBERGAGGTGCAVVUAAQDSNMWMSCEFLADNSEITATCAATPVSVUDY  81
      ||  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db     106  gllaevprlagidpdeiaadprla-----telreal-dvrgvrlkarpkissvld  155
QY      82  DQPGSTQAKLAVIVPRTGGRFG--ALLVNEGPRG--'ASAVDMAAMRAIADTDILKHFND  138
      ||  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db     156  ggrfglgaavvaditlqavstvayawwllsfgstksavgtlagnaupalltllleka  215
QY     139  LVGFDPRG---VGHSTRPLRCRTDAEFDAYRRDPMADYSRAGVTHVEQVYKOLADQCYD  194
      ||  ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db     216  slgtumgrldpseiralrcetesse-----tpaerpsaa-----lpglhalgnadtvl  265
QY     195  RMGSEFLANIGTASVARDMDMVRQALDDDDINTL-GVSYGTKLGTALNLFKFGHNVRAMVL  253
      ||  ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db     266  glglaf'agveeaaiaelylqv-'galganaalrlapghaf-----fvl  305
QY     254  DGAIDPAPVPIEESISOMAGFOTAFND---YAADCARSPACRPTGDSACQVNRHNLVDP  310
      ||  ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db     306  --glcpetaavagqlaashhgfriaeqdprnaatlcagskc-----asawmet-kmaer  357
QY     311  LVQ-----KPKK-----TSDPRGLSYADATVTGTINAL---YSP  340
      ||  ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db     358  lvetapelldgsltlvhhsgackgcarpkpsellvlgprsg-'yglvvnnganglpseaytd  415
QY     341  QRKKYTLTSGLLG--LQSGSDAGD  361
Db     416  engngsalarlgrlvrqnpkage  438

```

RESULT 12

ID W72061 standard; Protein; 732 AA.

AC W72061;

DT 07-DEC-1998 (first entry)

DE HSV-2 strain SB5 Contig ID 90 ORF#1 protein.

KW HSV-2 strain SB5; immunological response induction; therapy;

XX

XX

FT Misc-difference 667

FT	Misc-difference	691

```

FT      /note= "unspecified amino acid"
FT      Misc-difference 732
FT      /note= "encoded by C"
XX
XX      WO9820016-A1.
XX
XX      14-MAY-1998.
XX
XX      31-OCT-1997; 97WO-US20016.
XX
XX      09-JUN-1997; 97US-0049018.
XX      04-NOV-1996; 96US-0030279.
XX
XX      (SMIK ) SMITHKLINE BEECHAM CORP.
XX
XX      Chan JY, Dabrowski-Amaral CE, Delvecchio AM, Dillon SB;
XX      Esser KM, Leary JY;
XX      WPI; 1998-286847/25.
XX      N-PSDB; V62145.
XX
XX      Herpes simplex virus type-2 sequences - useful in, e.g. prevention
XX      and treatment of infection or inducing immunological response in
XX      mammal
XX
XX      Claim 10; Page 63; 748pp. English.
XX
XX      This sequence represents a Herpes simplex virus type-2 (HSV-2) protein
XX      sequence of the invention. This sequence was isolated from a HSV-2 strain
XX      SB5 (deposited as ATCC VR-2546) DNA fragment designated Contig ID 90.
XX      Based on homology, this sequence is a large tegument protein.
XX      The proteins can be used for the treatment or prevention of disease, to
XX      induce an immunological response in a mammal or to identify inhibitors,
XX      activators or novel antivirals. Antagonists of the proteins can be used
XX      to inhibit a viral polypeptide. The DNA sequence of a vector containing
XX      it can also be used to induce an immunological response in a mammal.
XX
XX      Sequence 732 AA:
XX
XX      Query Match 4.1%; Score 111; DB 19; Length 732;
XX      Best Local Similarity 21.2%; Pred. No. 0.22;
XX      Matches 103; Conservative 42; Mismatches 165; Indels 176; Gaps 23;
XX
XX      1 AAMRRRPSSALISGGLIGLLAAPPACATGEPGAGQTPGAVVAQDSWNSCRF 60
XX      358 aaahltlpaaaytandlylaavl-----gapvv-----alrnt 392
XX
XX      61 IADTSEIRFARCATVSVPVDDPGGTQAKLAVIRVPATGQRFGALLVMPGPGASAVM 120
XX      393 tafreselelclt-----fda-----fpggpdalrty 422
XX
XX      121 VAMAPAIADTDI-----LRHFDLVGFPFGVGHSTPALRCRTDAEPFAYRRPMADYS 174
XX      423 vs-----sdletwavgllhldlnpienaclaqlprlsa-----liaerpladgp 467
XX
XX      175 PA-----GTVHVEQY-----ROLAODCVDRMGESFLANIGTASVARDMD--M 215
XX      468 pclvlydlsmtpyavilwaeeppgyprfvgseateelpfvaiaqdvlaasaadapff 527
XX
XX      216 VFOALGDDQINLYGYSYGRKLTGATVLERFGTHFRAMVLDGIDPAVSPIEESISQMAGRQ 275
XX      528 aatall-----grpfdaalilgellfpghpyqyrpladag-- 561
XX
XX      276 TAFNDYAADCARSPACPLGTDSAQWVNRVHALVPLVQRPGKTSDRGLS--YADATGG 332
XX      562 ----psapkaardprladgsggspe-----dp-aapparqadgpyvlapltlatcg 610
XX
XX      333 TINALYSPQRKRLTSGLLGLGRGSDAGDLLVADDDYDGRDADGHSNDODAF----- 385
XX      611 ---epvprtmawvl-hgl-----eelaesedagpfcnpapalppatd 650
XX
XX      386 -NAVRCVVAPTPADPAWVAADQRIQVAPFLSYGQFTGS---APRDLCALMVPATSTP 441

```

```

DB      :      ||| | ||      : : : | ||      |||      : : ||
DB      651 qsvtsgypgrlpgrpax-----taretrpsvppgqtlgrvpapc-----xdprpsppp 700
XX
XX      442 HPAAPA 447
XX
XX      701 sppada 706
XX
XX      RESULT 13
XX      W22607
XX      ID W22607 standard; Protein; 1996 AA.
XX
XX      W22607;
XX
XX      27-FEB-1998 (first entry)
XX
XX      Platenolide synthase ORF2 protein.
XX
XX      Tyactone synthase gene cluster; tylG gene; multifunctional protein;
XX      polyketide; tyactone synthesis; antibiotic; tylosin.
XX
XX      Streptomyces ambotaciens.
XX
XX      Key Location/Qualifiers
XX      Modified-site 1
XX
XX      Domain
XX      /note= "encoded by GTC"
XX      15..418
XX      /note= "ketosynthase domain, KS'(s)"
XX      525..882
XX      /note= "acyltransferase domain, AT(s)"
XX      942..1025
XX      /note= "acyl carrier protein domain, ACP(s)"
XX      1060..1483
XX      /note= "ketosynthase domain, KS1"
XX      1596..1953
XX      /note= "acyltransferase domain, AT1"
XX      2232..2416
XX      /note= "ketoreductase domain, KR1"
XX      2533..2616
XX      /note= "acyl carrier protein domain, ACP1"
XX      2641..3064
XX      /note= "ketosynthase domain, KS2"
XX      3184..3520
XX      /note= "acyltransferase domain, AT2"
XX      3546..3727
XX      /note= "dehydratase domain, DH2"
XX      4083..4268
XX      /note= "ketoreductase domain, KR2"
XX      4374..4457
XX      /note= "acyl carrier protein domain, ACP2"
XX
XX      EP791655-A2.
XX
XX      27-AUG-1997.
XX
XX      19-FEB-1997; 97EP-0301056.
XX
XX      22-FEB-1996; 96US-0012078.
XX
XX      (ELIL ) LILLY & CO ELI.
XX
XX      Dehoff BS, Kuhstoss SA, Rosteck PR, Sutton KL;
XX      WPI; 1997-418046/39.
XX      N-PSDB; T80414.
XX
XX      DNA encoding Streptomyces fradiae tyactone synthase domain - for
XX      production of tylosin-related polyketide compounds
XX
XX      Example 2; Pages 149-155; 220pp; English.
XX
XX      W22606-W22610 represent proteins encoded by the platenolide synthase gene
XX      cluster. The gene cluster is also referred to as the smg gene, and was

```

CC isolated from Streptomyces ambifaciens. These sequences are
 CC multi-functional proteins which direct the synthesis of the polyketide
 CC platenolide. Platenolide is the basic building block of the macrolide
 CC antibiotic spiramycin. The DNA encoding this sequence was used along with
 CC the tyg gene (see T80413) to create a hybrid ORF1 sequence (see T80415).
 CC The tyg gene is the tylation synthase gene cluster of the invention.
 CC The tyg sequence was isolated from Streptomyces fradiae, and encodes
 CC multifunctional proteins which direct the synthesis of the polyketide
 CC tylation. Tylation is the basic building block of the antibiotic
 CC tylosin. The hybrid sequence can be used to transform S. ambifaciens
 CC lacking the smg ORF1 sequence, or S. fradiae lacking the tyg ORF1
 CC sequence, so that they can produce polyketides. The DNA sequence can be
 CC modified so as to alter the type of carboxylic acids incorporated, the
 CC number of carboxylic acids incorporated and/or the post-condensation
 CC reactions performed, thereby resulting in novel tylosin-related
 CC polyketides.

XX Sequence 1996 AA:

Query Match 4.0%; Score 108.5; DB 18; Length 1996;

Best Local Similarity 23.0%; Pred. No. 1.5; Mismatches 186; Indels 213; Gaps 32;

Matches 134; Conservative 50; Mismatches 186; Indels 213; Gaps 32;

QY 26 AAPPLAGATEEPGA-----GQTPG---APVVAPOQSMNSCREFIADTSEIRTA 71
 DB 1426 aappeggaadpagaavvgllrvvagaegpgrftlvdvgtqaslrallpglladagqsavr 1485
 QY 72 CATVSP--VDYDP-----GTOAKLAVIRP-ATGGRFGALLVNPGRGASAVDMVAAM 124
 DB 1486 dgrvtvprlrvpadpvrphggtaadgtgagpsatldegtvilt-99tga---laaet 1540
 QY 125 APAIADTILRHFDLVF---DPRGVH-----STPALRCRTDAEFVARR--DPM 170
 DB 1541 arlhvdtkvrhlllvgrigpdaqvdlvaeltsgaevavrcdvtldaltrllldai 1600
 QY 171 ADYSP-----AGV-----THVEQVYRQLAQCDCVDRMGF 198
 DB 1601 pdehplcvvhtagvlldgvltsagtaeridtlvrkdaavhldeltreigr--vplvly 1658
 QY 199 SFL-ANIGTASVA-----RDMWV---RQALGDDQINLYGSGTKLG-----TA 239
 DB 1659 ssvsaatlgsaggaagaaanafmdalaatrcaghpals-lygwswsgvlatgldgdaa 1717
 QY 240 YLEFFG-----THVRAMVLDGAI---DPAVSPIESISQMGFOTAFENDVADCARSPAC 291
 DB 1718 rvrtsglarpdagaaalldlalttrpeallrvlridlaaag-atalpevlridlaaypa- 1775
 DB 292 PLGTDSAQWVNRHVALVDPVQKPKGTSDPRLSYADATTGTINALXSPQRMYLTSGL 351
 DB 1776 -----darstfgaaagtg-----1788
 QY 352 GLQKGSAGDLVLADDDYDGRDADGHSNDQAFNAYRCVAPRPADPAVAAADQIRQ 411
 DB 1789 -----dedg-----avr--papapadaagflaa-----1809
 QY 412 VAPFLSYGQFTG--SAPRDLCALMPVATSTPRHARAAGAGVVVVS---THNDATPYQS 467
 DB 1810 -----tiagrsapertallldivrtlev---aavlgngpaaigaartfkdgfsilt 1858
 QY 468 GVDLARQLAGRLITFDGTQ--HFAVFDGNOCVDSAVNMHYFLDG 508
 DB 1859 avdlrnln-----trglrlpatlvfg--hptrlalaeillldg 1895
 RESULT 14
 W23717 ID W23717 standard; Protein; 1996 AA.
 XX W23717;
 XX 27-FEB-1998 (first entry)
 DT XX
 XX

DE Platenolide synthase ORF2 protein.

XX Platenolide synthase gene cluster; platenolide production; smg gene;
 KW multi-functional protein; macrolide antibiotic; spiramycin.

OS Streptomyces ambifaciens.

XX Key Location/Qualifiers

FT Domain 35..459 "ketosynthase domain, KS3"

FT Domain 582..933 /note="acyltransferase domain, AT3"

FT Domain 957..1155 /note="dehydratase domain, DH3"

FT Domain 1523..1707 /note="ketoreductase domain, KR3"

FT Domain 1812..1895 /note="acyl carrier protein domain, ACP3"

FT Domain EP791656-A2.

FT Domain 27-AUG-1997.

PE 19-FEB-1997; 97EP-0301066.

PR 22-FEB-1996; 96US-0012050.

PA (EIL) LILLY & CO E.LI.

PI Burgett SG, Kuhstoss SA, Rao RN, Richardson MA;

PI Rostock PR;

XX WPI: 1997-418047/39.

DR N-PSDB: T78508.

PS Claim 8; Pages 48-54; 81pp; English.

CC W23716-W23720 represent proteins encoded by the platenolide synthase gene

CC cluster of the invention. The gene cluster is also referred to as the

CC smg gene, and was isolated from Streptomyces ambifaciens. These

CC sequences are multi-functional proteins which direct the synthesis of the

CC polyketide platenolide. Platenolide is the basic building block of the

CC macrolide antibiotic spiramycin. The DNA can be used to produce compounds

CC exhibiting antibiotic activity based on the platenolide structure,

CC including specifically the macrolide antibiotic spiramycin and spiramycin

CC analogues and derivatives. Modifications of the platenolide synthase DNA

CC sequence can be made so as to change the number and type of carboxylic

CC acids incorporated into the growing polyketide chain and to change the

CC kind of post-condensation processing that is conducted.

XX Sequence 1996 AA:

Query Match 4.0%; Score 108.5; DB 18; Length 1996;

Best Local Similarity 23.0%; Pred. No. 1.5; Mismatches 186; Indels 213; Gaps 32;

Matches 134; Conservative 50; Mismatches 186; Indels 213; Gaps 32;

QY 26 AAPPLAGATEEPGA-----GQTPG---APVVAPOQSMNSCREFIADTSEIRTA 71
 DB 1426 aappeggaadpagaavvgllrvvagaegpgrftlvdvgtqaslrallpglladagqsavr 1485
 QY 72 CATVSP--VDYDP-----GTOAKLAVIRP-ATGGRFGALLVNPGRGASAVDMVAAM 124
 DB 1486 dgrvtvprlrvpadpvrphggtaadgtgagpsatldegtvilt-99tga---laaet 1540
 QY 125 APAIADTILRHFDLVF---DPRGVH-----STPALRCRTDAEFVARR--DPM 170
 DB 1541 arlhvdtkvrhlllvgrigpdaqvdlvaeltsgaevavrcdvtldaltrllldai 1600
 QY 171 ADYSP-----AGV-----THVEQVYRQLAQCDCVDRMGF 198


```

Db      1601 pdehplltcvvhtagvlddyvlsaqtaeridvtlirpkadaavhldel:trigr--vplivly 1658
Qy      199  sFL-ANIGTASVA-----RDMDMV-----ROALGDDDLINVLGYSYTKLG-----TA 239
Db      1659  svsaaatlgsaggaayaaanafmdaalaaricaaghpals-lygwwsyyglatlgidgaada 1717
Qy      240  YLERFG-----THVRAWYLDGAI--DPAVSPIEESISOMAGFOFANDYADCARSPAC 291
Db      1718  rrvrrsglapidagaalldlratlrtrpepal]pyrlldraag-atalpevrlidiagva- 1775
Qy      232  PLGTDSAOWNRHVALVPLVQKPKGTSIDPRGISTADATGTGINALVSPQRKMYLTSGL 351
Db      1776  -----darstpgaaagt-----1788
Qy      352  GLQKSDAGDLLVADDTYDGRDADGHSNDQDAFNVAKCVYAPTPADPAWYAADQRIK 411
Db      1789  -----dedg-----avr--papapadaagtlaa-----1809
Qy      412  VAPFLSYGQFTG-SAPRDLCALMPVPATSRPHPAAGAGKVVVVS-----THHDPRTPYQS 467
Db      1810  -----rlagrsapertaillldivrtvev--aavlgpbdpaalgaartlfxkdagsalt 1858
Qy      468  GVDLAROLGAPLRFDDGTQ--HTAAVFGNGNCVDSAAVNHFFELG 508
Db      1859  avdlinrn---crglrlipactlvtd--hpelrlaellailldg 1895

```

CC Products from the present invention can also be used for screening
CC compounds to identify agonists and antagonists of *E. faecalis* protein
CC activity.

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 12, 2001, 18:01:32 ; Search time 3358.72 Seconds
(Without alignments)
1883.980 Million cell updates/sec

Title: US-09-461-774-5
Perfect score: 903
Sequence: 1 gattacgaaagagctggyga.....aggagaggtgaaagcttaa 903

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 7991742 seqs, 3503743858 residues
Total number of hits satisfying chosen parameters: 15983484

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*

1:	gb_est1:*
2:	gb_est2:*
3:	gb_est3:*
4:	gb_est4:*
5:	gb_est5:*
6:	gb_est6:*
7:	gb_est7:*
8:	gb_est8:*
9:	gb_est9:*
10:	gb_est10:*
11:	gb_est11:*
12:	gb_est12:*
13:	gb_est13:*
14:	gb_est14:*
15:	gb_est15:*
16:	gb_est16:*
17:	gb_est17:*
18:	gb_est18:*
19:	gb_est19:*
20:	gb_est20:*
21:	gb_est21:*
22:	gb_est22:*
23:	gb_est23:*
24:	gb_est24:*
25:	gb_est25:*
26:	gb_est26:*
27:	gb_est27:*
28:	gb_est28:*
29:	gb_est29:*
30:	gb_est30:*
31:	gb_est31:*
32:	gb_est32:*
33:	gb_est33:*
34:	gb_est34:*
35:	gb_est35:*
36:	gb_est36:*
37:	gb_est37:*
38:	gb_est38:*
39:	gb_est39:*
40:	gb_est40:*
41:	em_est41:*
42:	em_estfun:*
43:	em_esthum1:*

44:	em_esthum2:*
45:	em_esthum3:*
46:	em_esthum4:*
47:	em_esthum5:*
48:	em_esthum6:*
49:	em_esthum7:*
50:	em_esthum8:*
51:	em_esthum9:*
52:	em_esthum10:*
53:	em_esthum11:*
54:	em_esthum12:*
55:	em_esthum13:*
56:	em_esthum14:*
57:	em_esthum15:*
58:	em_esthum16:*
59:	em_esthum17:*
60:	em_esthum18:*
61:	em_esthum19:*
62:	em_esthum20:*
63:	em_estin1:*
64:	em_estin2:*
65:	em_estin3:*
66:	em_estin4:*
67:	em_estov1:*
68:	em_estov2:*
69:	em_estp11:*
70:	em_estp12:*
71:	em_estp13:*
72:	em_estp14:*
73:	em_estp15:*
74:	em_estro1:*
75:	em_estro2:*
76:	em_estro3:*
77:	em_estro4:*
78:	em_estro5:*
79:	em_estro6:*
80:	em_estro7:*
81:	em_estro8:*
82:	em_estro9:*
83:	em_estro10:*
84:	em_estro11:*
85:	em_estro12:*
86:	em_estro13:*
87:	gb_est41:*
88:	gb_est42:*
89:	gb_est43:*
90:	gb_est44:*
91:	gb_est45:*
92:	gb_est46:*
93:	gb_est47:*
94:	gb_est48:*
95:	gb_est49:*
96:	gb_est50:*
97:	gb_est51:*
98:	gb_est52:*
99:	gb_est53:*
100:	gb_est54:*
101:	gb_est55:*
102:	gb_est56:*
103:	gb_est57:*
104:	gb_est58:*
105:	gb_est59:*
106:	gb_est60:*
107:	gb_est61:*
108:	gb_est62:*
109:	gb_est63:*
110:	gb_est64:*
111:	gb_est65:*
112:	em_esthum21:*
113:	em_esthum22:*
114:	em_esthum23:*
115:	em_estom1:*
116:	em_estom2:*

117: em_estp16:*
 118: em_estp17:*
 119: em_estp18:*
 120: em_estp19:*
 121: em_estp20:*
 122: em_estp21:*
 123: em_estp22:*
 124: em_estp23:*
 125: em_estp24:*
 126: em_estp25:*
 127: em_estp26:*
 128: em_estp27:*
 129: em_estp28:*
 130: em_estp29:*
 131: em_estp30:*
 132: em_estp31:*
 133: em_estp32:*
 134: em_estp33:*
 135: em_estp34:*
 136: em_estp35:*
 137: em_estp36:*
 138: em_estp37:*
 139: em_estp38:*
 140: em_estp39:*
 141: em_estp40:*
 142: em_estp41:*
 143: em_estp42:*
 144: em_estp43:*
 145: em_estp44:*
 146: em_estp45:*
 147: em_estp46:*
 148: em_estp47:*
 149: em_estp48:*
 150: em_estp49:*
 151: em_estp50:*
 152: em_estp51:*
 153: em_estp52:*
 154: em_estp53:*
 155: em_estp54:*
 156: em_estp55:*
 157: em_estp56:*
 158: em_estp57:*
 159: em_estp58:*
 160: em_estp59:*
 161: em_estp60:*
 162: em_estp61:*
 163: em_estp62:*
 164: em_estp63:*
 165: em_estp64:*
 166: em_estp65:*
 167: em_estp66:*
 168: em_estp67:*
 169: em_estp68:*
 170: em_estp69:*
 171: em_estp70:*
 172: em_estp71:*
 173: em_estp72:*
 174: em_estp73:*
 175: em_estp74:*
 176: em_estp75:*
 177: em_estp76:*
 178: em_estp77:*
 179: em_estp78:*
 180: em_estp79:*
 181: em_estp80:*
 182: em_estp81:*
 183: em_estp82:*
 184: em_estp83:*
 185: em_estp84:*
 186: em_estp85:*
 187: em_estp86:*
 188: em_estp87:*
 189: em_estp88:*

190: gb_gss25:*
 191: gb_gss26:*
 192: gb_gss27:*
 193: gb_gss28:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	78.8	8.7	340	7	AA465061
2	43.2	4.8	430	193	AL390626 Leishman
3	40.6	4.5	925	190	AL053013 Drosophila
4	37.6	4.2	574	97	AW982729 HVSME900
5	37.6	4.2	1101	190	AL108460 Drosophila
6	36.6	4.1	249	140	F14761 SSC908 Por
7	36.6	4.1	477	108	BE417569 MGC022.D0
8	36.6	4.1	478	107	BE415878 MGC002.A0
9	36.6	4.1	484	108	BE416420 MGC008.E0
10	36.6	4.1	849	97	AW926475 HVSME900
11	36.2	4.0	673	29	AU090557 AU090557
12	36	4.0	416	108	BE416649 MGC009.H0
13	36	4.0	439	171	A0846629 LMAJFV1.1
14	36	4.0	465	171	A0851993 LMAJFV1.1
15	36	4.0	467	110	BE605099 WHE1701-1
16	36	4.0	600	107	BE415989 MGC003.B0
17	36	4.0	1159	190	AL106041 Drosophila
18	35.6	3.9	504	136	BE859829 UI-M-AHO-
19	35.6	3.9	607	110	BE586256 WHE501.B0
20	35.6	3.9	1063	192	AL273193 Tetraodon
21	35.4	3.9	539	93	AM672855 base08.y
22	35.4	3.9	584	107	BE388729 BE388729
23	35.4	3.9	689	137	BE905873 BE905873
24	35.2	3.9	543	29	AU090556 AU090556
25	35.2	3.9	910	190	AL065629 Drosophila
26	35	3.9	340	108	BE417796 BE417796
27	35	3.9	403	110	BE606633 WHE0901.G
28	35	3.9	435	109	BE517717 WHE0802.F
29	35	3.9	438	107	BE400743 AMB007.B0
30	35	3.9	440	108	BE417801 MGC024.G0
31	35	3.9	442	108	BE416187 MGC005.E1
32	35	3.9	452	108	BE426473 WHE0335.C
33	35	3.9	475	97	AW983152 HVSME900
34	35	3.9	483	108	BE415589 MGC009.C0
35	35	3.9	584	107	BE405684 WHE1214.B
36	35	3.9	816	105	BE188063 CFC341-R
37	34.8	3.9	395	108	BE417757 MGC024.B1
38	34.8	3.9	1054	193	AL341522 Tetraodon
39	34.6	3.8	661	107	BE405558 WHE1212.G
40	34.4	3.8	542	107	BE415790 MGC001.A1
41	34.4	3.8	623	23	AL637198 603002B11
42	34.4	3.8	659	107	BE415854 MGC001.G0
43	34.4	3.8	674	107	BE416059 MGC004.C0
44	34.4	3.8	1020	190	AL067513 Drosophila
45	34.2	3.8	503	105	BE215761 HV_CBD000

ALIGNMENTS

RESULT 1
 AA465061 340 bp mRNA
 LOCUS EST_Altonzo Mtb FOX (5% oxygen) expression library Mycobacterium
 DEFINITION tuberculosis CDNA clone Altonzo similar to Acyl-CoA ligase, mRNA
 ACCESSION AA465061
 VERSION AA465061.1 GI:2191201
 KEYWORDS EST.

SOURCE	Mycobacterium tuberculosis.
ORGANISM	Mycobacterium tuberculosis Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium 1 (bases 1 to 340)
REFERENCE	Imboden, P. and Schoolnik, G. K.
TITLE	Construction and characterization of a partial Mycobacterium tuberculosis cDNA library of genes expressed at reduced oxygen tension
JOURNAL	Gene 213 (1-2), 107-117 (1998)
MEDLINE	98296259
COMMENT	Contact: Paul Imboden

FEATURES

source	Email: plimbodengimm.unibe.ch see also Genbank Accession Number U64610; homology to Safirmaycin synthetase Insert Length: 340 Std Error: 10.00 Plate: 1 row: 8 column: 6. Location/Qualifiers 1..340
--------	--

```

/organism="Mycobacterium tuberculosis"
/strain="PI, ATCC 51910"
/db_xref="taxon:1773"
/clone="Alonzo"
/clone_lib="Mt.b. FOX (5% oxygen) expression library"
/dev_stage="late log phase"
/lab_host="Escherichia coli strain Novablue (Novagen)"
/notes="Vector: pET24a (+) (Novagen); Site_1: EcoRI;
Site_2: EcoRI; total RNA isolated from liquid cultures
grown to late log phase in Middlebrook 7H9/ADC in an
atmosphere of 5% oxygen, 5% CO2 and 90% Nitrogen. Random
fragments were amplified from total RNA by PCR with the
primer UPN6 (5'-GCCGAGCTCTCCAGATCCNNNNN-3') and cloned
into EcoRI site of pET24a (+). Clones from M.tb ribosomal
RNA were detected by hybridization and removed."

```

Query Match	8.7%	Score 78.8	DB 7	Length 340
Best Local Similarity	54.9%	Pred. No. 2.1e-12		
Matches 180; Conservative	0	Mismatches 142	Indels 6	Gaps 1

278	caagctcgcgcgttggaacgactcgtcgaacatgctgtgcgcgccgaggaatccccgc	337
Db	19 CACGGTCGGGTGAGGCCGATTAACGAATAAGCAGCCAGCCACAGCGGCAAGCTGCCC	78
OY	338 catcaattatcgaagcttgatctgctcgaatcgtacgtcgcctcggaatggatataccttcaag	397
Db	79 CGGTGCTGTTGAGGCTGATGTGCTTATTTGGACTCGCCGCGACACATGCTCCGGCTTTCT	138
OY	398 aagaacgaatcacatctaccgcgcatcttgcaatacacctccggtgtcacccgcgaacgcgcg	457
Db	139 CTCGTCAACACACCGGGGGCGCTTATCTCCAATACAGTCCGGATGCACCGCGTACGCCGG	198
OY	458 ctgcgcgttggtgatatgcccatcagaacgltcgggttaatttcogaacagctgactgtcgtcgt	517
Db	199 CCGGAGTCATATTGTGTCCACACGAGATGTCAATGCCAATGTGACACCAAGATGTACGGCT	258
OY	518 acttcgcgataccgaacgagatctccacgcgcgaattccgcactgtatctcgtctaacct	577
Db	259 ATTTCGGGGGATCCCGCAAGAATTCGACACCGGGA-----CTGTGTGTCTGTGGCTGCTT	312
OY	578 tctaacacgacatggtgttgtaataag	605
Db	313 TGTATCACGATATGGGCTGTGTTCTCGG	340

RESULT	2
P929R	
LOCUS	P929R 430 bp DNA
DEFINITION	GSS 25-Jul-2000
	Leishmania major Fried in PAC P929 right end-sequence, similar to
	FR:087548 OR87548 ACSA (FRAGMENT). [0] +3. . . , N=164, Prob=2.5e-14,
	genomic survey sequence.
ACCESSION	AF390626 AF390626 GI:9501602
VERSION	GSS.
KEYWORDS	
SOURCE	Leishmania major.
ORGANISM	Leishmania major

REFERENCE	1 (bases 1 to 430)
AUTHORS	Ivens, A.C., Lewis, S.M., Bagherzadeh, A., Zhang, L., Chan, H.M. and Smith, D.F.
TITLE	A physical map of the <i>Leishmania major</i> Friedlin genome
JOURNAL	Genome Res. 8 (2), 135-145 (1998)
MEDLINE	98146435
REFERENCE	2 (bases 1 to 430)
AUTHORS	Taylor, R.G., Huckle, E.E.J., Ivens, A.C., Rajandream, M.A. and Barrell, B.G.
TITLE	Direct Submission
JOURNAL	Submitted (24-JUL-2000) <i>Leishmania major</i> Friedlin genome sequencing project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA. E-mail: barrell@sanger.ac.uk and allicat@sanger.ac.uk
COMMENT	see http://www.ebi.ac.uk/parasites/leish.html

FEATURES	Location/Qualifiers
source	1. .430

```

/organism="Leishmania major"
/strain="Friedlin"
/db_xref="taxon:5664"
/clone="PAC P929"
BASE COUNT      61 a    141 c    142 g    86 t
ORIGIN

```

Query Match	4.8%	Score	43.2	DB	133	length	430
Best Local Similarity	61.6%	Pred. No.	0.051				
Matches	69	Conservative	0	Mismatches	43	Indels	0
						Gaps	0

Qy 166 ggcgtgtcagagccggagcgatcgccgtgtcgcgctttcgtgttcccaagaagcgagc 217
Db 312 gcttaccacagcttccggcgatcgccatgtccggcgtgtcgtatgacttmttgagcccg 363

RESULT	3	
CNS0091P		
LOCUS		
DEFINITION	CNS0091P 925 bp DNA GSS 03-JUN-1999	
ACCESSION	Drosophila melanogaster genome survey sequence TE13 end of BAC #	
VERSION	BAR19D16 of Rp1-98 library from Drosophila melanogaster (fruit	
KEYWORDS	fly), genomic survey sequence.	
SOURCE	AL053013	
ORGANISM	AL053013.1 GI:4934461	
	GSS.	
	fruit fly.	
	Drosophila melanogaster	
	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;	
	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;	
	Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.	
	1 (bases 1 to 925)	
	Genoscope.	
REFERENCE	Direct Submission	
AUTHORS		
TITLE		
JOURNAL	Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :	

and Genevieve Payan. It has been constructed in the vector pBelBAC11.

FEATURES
source

1..1101
/organism="Drosophila melanogaster"
/plasmid="pBelBAC11"
/db_xref="taxon:7227"
/clone_1lb="DrosBAC"
/note="BACN371.08"
/note="end : SP6"

BASE COUNT 254 a 176 c 160 g 152 t 359 others
ORIGIN

Query Match 4.2%; Score 37.6; DB 190; Length 1101;
Best Local Similarity 13.4%; Pred. No. 2.6;
Matches 41; Conservative 143; Mismatches 122; Indels 0; Gaps 0;

QY 35 cccctgacgcgtgcacgtatcgcgcgaacgtgaatgtgacagagcgtgagccgtt 94
1048 SBBSSGSSSSBSSBSSSTSSSSSTSSSSSTSSBTTSTSSSTTTTSTTTTSTTB 989
QY 95 gtgttccacaggtgacgcgtgtgtatctctgcgcagaggaactgcagtcgtcg 154
988 TBBSSBSSBSSSTSSSBTTSKBTTSSBSSBSSSTSSASBSSSSSSBTTBTTBBS 929
QY 155 cctatctgcgcgtgtgcagccgcgcgcgtgcgcgttgcgttgcacaaagcg 214
928 TSSSSGSSSSBSSBSSBTTTBTSTSTSSSTSSSTSTSSBSSSSSSSSBSS 869
QY 215 gcgttaccgatgaacgttcgttcgttgcgttcgttcgttcgttcgttcgttc 274
868 SBTSTTSSBSSBSSSTSSBTTCTSKCTBTSTBTSTBSSBSSSTSTSSGSSB 809
QY 275 ctacatcgtctgcgtgcgcgcgttcgttcacatgttcgcgcgcgcgcgcgcgc 334
808 SCGSTRGSSGSCBCKSTSGSTRSGTTCGBCSSTGCCGCCCTCCTCSTGTCGS 749
QY 335 cgcacat 340
Db 748 CSCSGT 743

RESULT 6

LOCUS F14761 249 bp mRNA EST 09-SEP-1996
DEFINITION SSC9C08 Porcine small intestine cDNA library Sus scrofa cDNA clone
c9c08, mRNA sequence.
F14761
VERSION F14761.1 GI:972340

KEYWORDS EST.
SOURCE pig.
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.

REFERENCE 1 (bases 1 to 249)
AUTHORS Wintero, A.K., Fredholm, M. and Davies, W.
TITLE Evaluation and characterization of a porcine small intestine cDNA
library: analysis of 839 clones
JOURNAL Mamm. Genome 7 (7), 509-517 (1996)
MEDLINE 96372607

COMMENT Contact: A.K. Wintero
Department of Animal Science and Animal Health, Division of Animal
Genetics, The Royal Veterinary and Agricultural University
Bulowsvej 13, 1870 Frederiksberg C, Denmark.

FEATURES
source

1..249
/organism="Sus scrofa"
/db_xref="taxon:9623"
/clone_1lb="C9c08"
/note="directionally cloned cDNA in xli-blue MRF"

BASE COUNT 37 a 75 c 73 g 58 t 6 others

ORIGIN

Query Match 4.1%; Score 36.6; DB 140; Length 249;
Best Local Similarity 61.3%; Pred. No. 3.9;
Matches 57; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 103 acgggtgacgcgtgtgtatctctctcgcagaggaactgaatcgtctgcctatctc 162
Db 157 AAGGGGTGCGGTGTGTGACGTGTCTCGCGAGGAGATGTGACAGAGCGCCCTGCTC 216

QY 163 ggcgcgtgcagc 195
Db 217 CGNCCCTCCACATCGGNCATGCGCTGAGGCG 249

RESULT 7

LOCUS BEA17569/c 477 bp mRNA EST 24-JUN-2000
DEFINITION MUG022.D04R90620 TREC MUG Wheat Spikelet Library Triticum aestivum
cDNA clone MUG022.D04, mRNA sequence.
BEA17569

ACCESSION BEA17569.1 GI:9415415
VERSION BEA17569
KEYWORDS EST.
SOURCE bread wheat.
ORGANISM Triticum aestivum

Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; Liliopsida; Poales; Poaceae; Triticum.
1 (bases 1 to 477)
Anderson, O.A., Appels, R., Bailey, P., Blake, T., Close, T., Cloutier,
S., Dubcovsky, J., Feuillet, C., Gale, M., Graner, A., Gustafson, P.,
Herrmann, R.G., Holton, T., Jacquemart, J.M., Jia, J., Jondrier, P.,
Langridge, P., Lazo, G.R., Lin, J.J., McGuire, P., Ogihara, Y.,
Pecchioni, N., Qualset, C., Schuch, W., Selvaraj, G., Shariflou, M.,
Sorrells, M., Warburton, M. and Wenzel, G.
International Triticale EST Cooperative (ITREC): Production of
Expressed Sequence Tags for Species of the Triticaceae
Unpublished (2000)

JOURNAL

COMMENT Contact: Ogihara Y
Kihara Institute for Biological Research, Yokohama City University
Makoka-cho 641-12, Totsuka-ku, Yokohama 244-0813, JAPAN
Tel: 81 45 820 1903
Fax: 81 45 820 1901
Email: ogihara@yokohama-cu.ac.jp
International Triticale EST Cooperative (ITREC)
http://wheat.pw.usda.gov/genome.

FEATURES
source

1..477
/organism="Triticum aestivum"
/cultivar="Norin 26"
/db_xref="taxon:4565"
/clone_1lb="MUG022.D04"
/clone_1lb="ITREC MUG Wheat Spikelet Library"
/tissue_type="young spikelets"
/dev_stage="Peaks' scale 6-7"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; M13 Reverse sequencing primer used. 1.2 kbp average
insert size."

BASE COUNT 103 a 140 c 134 g 100 t

ORIGIN

Query Match 4.1%; Score 36.6; DB 108; Length 477;
Best Local Similarity 49.7%; Pred. No. 4.4;
Matches 93; Conservative 0; Mismatches 94; Indels 0; Gaps 0;

QY 43 cgttcgcactgtatcggagaaactggaatgtgcacaggaactgagccgtgtgtcttc 102
Db 221 GTGGCCCTCTCTCTCCGCTCTTGTGGGAGACACGAGGAGTGTGGGATGAC 162

QY 103 acgggtgacgcgtgtgtgtatctctcgcagaggaactgaatcgtctgcctatctc 162
Db 161 GCCGCGTGCAGATGTCACGCCGCCACGACTTTCGCCAGACTCTCTCTCTCCGAC 102

QY	163	gagcggttgcgaagcgagcgatccgcttgcgcttgcgcttgcacaaagcgagcttacc	222
Db	101	GGCGACGACGACGAGTGGCGCGGATGATGCGCTCTTCTTGTGTGCTTGGCGGCTTGCC	42
QY	223	gatgaac	229
Db	41	GGCGAGC	35
RESULT	8		
BE415878/c			
LOCUS	BE415878	478 bp	mRNA
DEFINITION	MUG002.A103990520 ITREC MUG Wheat Spiklet Library Triticum aestivum		EST
ACCESSION	CDNA clone MUG002.A03, mRNA sequence.		24-JUL-2000
VERSION	BE415878		
KEYWORDS	BE415878.1	GI:9413724	
ORGANISM	EST.		
REFERENCE	bread wheat.		
AUTHORS	Triticum aestivum		
	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;		
	Magnoliophyta; Liliopsida; Poales; Poaceae; Triticum.		
	1 (bases 1 to 478)		
	Anderson,O.A., Appels,R., Bailey,P., Blake,T., Close,T., Cloutier		
	,S., Dubcovsky,D., Feuillet,C., Gale,M., Graner,A., Gustafson,P.,		
	Herrmann,R.G., Holton,T., Jacquemin,J.M., Jia,J., Joudrier,P.,		
	Langridge,P., Lazo,G.R., Lin,J.J., McGuire,P., Ogihara,Y.,		
	Pechion,L.N., Qualset,C., Schuch,K., Selvaraj,G., Shariflou,M.,		
	Sorrells,M., Warburton,M. and Wenzel,G.		
	International Triticale EST Cooperative (ITREC): Production of		
	Expressed Sequence Tags for Species of the Triticace		
	Unpublished (2000)		
JOURNAL	Contact: Ogihara Y		
COMMENT	Kihara Institute for Biological Research, Yokohama City University		
	Maoka-cho 641-12, Totsuka-ku, Yokohama 244-0813, JAPAN		
	Tel: 81 45 820 1903		
	Fax: 81 45 820 1901		
	Email: ogihara@yokohama-cu.ac.jp		
	International Triticale EST Cooperative (ITREC)		
	http://wheat.pw.usda.gov/genome.		
FEATURES			
Source	Location/Qualifiers		
	1..478		
	/organism="Triticum aestivum"		
	/cultivar="Norin 26"		
	/db_xref="taxon:4565"		
	/clone="MUG002.A03"		
	/clone_lib="ITREC MUG Wheat Spiklet Library"		
	/tissue_type="young spikelets"		
	/dev_stage="Feekes' scale 6-7"		
	/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:		
	XhoI; M13 Reverse sequencing primer used. 1.2 Kbp average		
	insert size."		
BASE COUNT	104 a	140 c	134 g
ORIGIN			100 t
Query Match	4.1%;	Score 36.6;	DB 107; Length 478;
Best Local Similarity	49.7%;	Pred. No. 4.4;	
Matches	93;	Conservative 0;	Mismatches 94; Indels 0; Gaps 0;
QY	43	cggttcgacgtgtatcgcgcgaagcgtgatgtgacacagagagctgagccggtgtgtgtcc	102
Db	221	CTGGCCCTCTCTCTCCGCGCTTCTTGTGGGAGCAGCGGGGTGATCTTGGGATGAC	162
QY	103	acgagtgacgggtgtgtatctctctcgcagaggaaccgagtaagtcgtgcgatctc	162
Db	161	GGCGCGGTGCGGATGATGACCGCGCAGCAGCTTGCACGACTCTCTGCTTCCGGAC	102
QY	163	gagcggttgcgaagcgagcgatccgcttgcgcttgcgcttgcacaaagcgagcttacc	222
Db	101	GGCGACGACGACGAGTGGCGCGGATGATGCGCTCTTCTTGTGTGCTTGGCGGCTTGCC	42
QY	223	gatgaac	229

DB	41	GCGCAGC 35
RESULT 9		
LOCUS	BEA16420/c	
DEFINITION	BEA16420 484 bp mRNA EST 24-JUL-2000	
ACCESSION	CDD0008.E04R90628 ITEC MUG Wheat Spikelet Library Triticum aestivum	
VERSION	MUG clone MUG008.E04, mRNA sequence.	
KEYWORDS	BEA16420.1 GI:9414266	
SOURCE	EST.	
ORGANISM	bread wheat. Triticum aestivum	
REFERENCE	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Triticum.	
AUTHORS	1 (bases 1 to 484) Anderson,O.A., Appels,R., Bailey,P., Blake,T., Close,T., Cloutier S., Dubcovsky,J., Feuillet,C., Gale,M., Graner,A., Gustafson,P., Herrmann,R.G., Holton,T., Jacquemin,J.M., Jia,J., Jouderier,P., Langridge,P., Lazo,G.R., Lin,J.J., McGuire,P., Ogihara,Y., Pechion,L.N., Qualset,C., Schuch,W., Selvaraj,G., Sharifou,M., Sorrells,M., Warburton,M. and Wenzel,G. Sotterelli,M., International Triticale EST Cooperative (ITEC): Production of Expressed Sequence Tags for Species of the Triticaceae Unpublished (2000)	
TITLE	Contact: Ogihara Y Kihara Institute for Biological Research, Yokohama City University Maoka-cho 641-12, Totsuka-ku, Yokohama 244-0813, JAPAN Tel: 81 45 820 1903 Fax: 81 45 820 1901 Email: ogihara@yokohama-cu.ac.jp International Triticale EST Cooperative (ITEC) http://wheat.pw.usda.gov/genome.location/Qualifiers	
JOURNAL		
COMMENT	1..484 /organism="Triticum aestivum" /cultivar="Norin 26" /db_xref="taxon:4565" /clone_id="ITEC MUG Wheat Spikelet Library" /tissue_type="young spikelets" /dev_stage="Peekes' scale 6-7" /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2: XhoI; M13 Reverse sequencing primer used. 1.2 kbp average insert size."	
FEATURES		
source		
BASE COUNT	105 a 138 c 137 g 102 t 2 others	
ORIGIN		
Query Match	4.1%, Score 36.6; DB 108; Length 484;	
Best Local Similarity	49.7%; Pred. No. 4.4;	
Matches	93; Conservative 0; Mismatches 94; Indels 0; Gaps 0;	
OY	43 cggtgcagcgtgatccgacgaactgtaatlgtagcacggaggctgagcgcttggttc 102	
Db	229 CTGGCCCTTCCTTCCTCCGCCGCTTTTGAGGAGCAACGCGGTGATCTTGGAATAC 170	
OY	103 acggatgacgcgctgtgtatctctgctccgcggaggaactcgaatcgtctgcctatcc 162	
Db	169 GCCGCCGTCGGGAGATGATGACCGCCGCACGACAGCTTGGCGAACTCTCTGCTCCGAGC 110	
OY	163 ggccgctgtcagagccgggacgcatcgccglccgcccttcgtttccaaagaagcggtacc 222	
Db	109 GGCGAGCAGCAGCATGGCGCGGAGATGATCGGCGCTTCTTGTGTGCTTGGCGGCGCTTCC 50	
OY	223 gatgaac 229	
Db	49 GCGCAGC 43	

Washtu leishmania Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810

Email: est@watson.wustl.edu
Library construction: Natalia S. Akopyants, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
If using this information please cite:

N.S. Akopyants and S.M. Beverley 'A survey of the *Leishmania* major Friedlin strain VI genome by shotgun sequencing' and the Washington University Genome Sequencing Center For information on obtaining clone material please contact: Natalia S. Akopyants Ph.D. (natalia@orcim.wustl.edu) and/or Stephen M. Beverley Ph.D. (beverley@wustl.edu)

(beverleyeborsam.wustl.edu)
Seq primer: -40RP from Gibco
Class: shotgun
High quality sequence stop: 389.

FEATURES	source	location/Qualifiers
		1..465
		/organism="Leishmania major"
		/strain="Friedlin strain V1"
		/db_xref="taxon:5664"
		/clone="LMAJFV1.im59g03"
		/clone_lib="Leishmania major FV1 random genomic library"
		/lab_host="TOP10 (Invitrogen)"
		/note="Vector: pZero-2 (Invitrogen); Site_1: EcoRV; Genomic DNA was isolated from stationary phase cells. For this library, DNA was sheared to give a tight size distribution of 1-1.5kb fragments, blunt-ended with T4 DNA polymerase, dephosphorylated with Shrimp Alkaline Phosphatase and ligated into pZero-2 vector's EcoRV site."
BASE COUNT		99 a 163 c 129 g 74 t

Query Match	4.0%;	Score 36;	DB 171;	Length 465;
Best Local Similarity	47.7%;	Pred. No. 6.6;		
Matches 105; Conservative	0;	Mismatches 115;	Indels 0;	Gaps 0;

Oy	47	cgagcgtctacccgcgaacgtgtgaatgtagcaagagcttbaacgcgtttgtgttccaa	106
Db	277	ctcactcttctccgccgacccctttagccgcactccgcttagccgctgtcttccgctc	218
Oy	107	gtgaaccgcgtgtgtagtctctgtctccgcagaagactcgatctacgtctgcctatctgcg	166
Db	217	gcctgccttgcgcgtctgtctctgtgacgacactcgtgtgattgaagtgcgaacctgcgaaggaag	158
Oy	167	cgttgcaagcccgagcgcacatcgccgttgcgccttctgltccaaagcgcgttacgcatg	226
Db	157	ggcagggcgaagggcgaccctctgattgacagagacgctttggccgaatgctggagctacacga	98
Oy	227	aacgttcgcattcgtaactgtatgtatctgttcgcggtgagc	266
Db	97	gatgctcacaactcgtttgtgacccgctcccgcgaggtctc	58

RESULT	15
LOCUS	BE605099/c
DEFINITION	BE605099 467 bp mRNA EST 21-Aug-2000 WHE1701-1704_A09.A09s2s wheat heat stressed spike cDNA library
ACCESSION	BE605099 BE605099
VERSION	BE605099.1
KEYWORDS	GI:3862369
SOURCE	EST.
ORGANISM	bread wheat. Triticum aestivum
REFERENCE	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Triticum. 1 (bases 1 to 467)
AUTHORS	Anderson,O.D., Chao,S., Choi,D.W., Close,T.J., Fenton,R.D., Han P.S., Hsia,C.C., Kang,Y., Lazo,G.R., Malatrasi,M., Miller,R., Nguyen,H.T., Rausch,C.J., Seaton,C.L., Tong,J.C. and Zhang,D.

TITLE	The structure and function of the expressed portion of the wheat
JOURNAL	genomes - Heat stressed spike cDNA library
COMMENT	Unpublished (2000)
	Contact: Olin Anderson

COMMENT
Contact: Olin Anderson
US Department of Agriculture, Agriculture Research Service, Pacific
West Area, Western Regional Research Center
800 Buchanan Street, Albany, CA 94710, USA
Tel: 5105595773
Fax: 5105595818

Email: Oanderson@pw.usda.gov
Sequence have been trimmed to remove vector sequence and low quality sequence with phred score less than 20
Seq primer: Stratagene SK primer.

FEATURES	
SOURCE	
1. .467 Location/Qualifiers /organism="Triticum aestivum" /cultivar="Chinese Spring" /db_xref="taxon:4565" /clone="WHEI701-1704_A09-A09" /tissue_type="Whole spike" /dev_stage="Spikes at 5, 10, 15 and 20 days after anthesis" /lab_host="E. coli SOLR" /note="Vector: Lambda Uni-ZAP XR, excised phagmid: Site_1: EcoRI; Site_2: XhoI; Spikes at 5, 10, 15 and 20 days after anthesis were heat stressed under two conditions at Texas Tech university (D. Zhang in RT Nguyenrith lab): (1) at 38 C for 4 hours and (2) 5 days of cyclic treatment of 38 C for 4 hours. Total RNA and poly(A) RNA were prepared, a cDNA library was made, and the cDNA clones were in vivo excised to give plusscript phagemids in the Tj Close lab (Choi, Close, Fenton, Malatrasi) at the University of California, Riverside. Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (all other authors)."	BASE COUNT ORIGIN 88 a 159 c 141 g 79 t

Query Match	4.0%	Score 36;	DB 110;	Length 467;
Best Local Similarity	48.5%	Pred. NO. 6.6;		
Matches 99; Conservative	0;	Mismatches 105;	Indels 0;	Gaps 0;

Oy	26	tttgcgataaccctbaacgagcggtcgaagctcgtatcgagacgtaatgtgcaaaagagc	85
Db	335	TTGGAGACACTTGGGGCTCTTCCCTCTTCTCCGGCGGTCTCTTGGGAGAGACACCGGG	299
Oy	86	tgaagccgcttgcgtgtccacagagttaccgcgctgcgtgtaatctctgcctcgaaggaactcagt	145
Db	295	TTGATCTTTCGGGATACAGCCACACGCTGGCGATGTGAACGCGCGGCACACAGCTTTCGAGC	238
Oy	146	acgttcctgcctatctcgtgcgctgtgaagagcagagcgcacgcctgcgtgcgcttcgtgtc	205
Db	235	TTCCTCGTCTCTCTGACGCGGCAGACAGAGGATGAGGCGGCGATGATGCGGCTTCTTCTGTC	176
Oy	206	cacaagcgcgcttaccgaatgaac	229
Db	175	TTCCTTGGCGGCGCTTCCCGGCGAGC	152

Search completed: January 12, 2001, 18:01:37
Job time: 21233 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 12, 2001, 20:12:43 : Search time 153.29 Seconds
(without alignments)
949.362 Million cell updates/sec

Title: US-09-461-774-5
Perfect score: 903
Sequence: 1 gattacgaagagctgga.....aggagaggtgaaagcttaa 903

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 280836 seqs, 80580151 residues

Minimum number of hits satisfying chosen parameters: 561672

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: Issued Patents_NA:*
2: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6.COMB.seq:*
4: /cgn2_6/ptodata/1/ina/PCTUS.COMB.seq:*
5: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	40	4.4	30001	1	US-08-125-468-1
2	40	4.4	30001	2	US-08-474-933-1
3	37	4.1	1899	3	US-09-354-129-7
4	35.4	3.9	11601	2	US-08-222-617A-3
5	35.4	3.9	11601	2	US-08-222-617A-24
6	34.8	3.9	38506	3	US-09-320-878-19
7	33.2	3.7	2219	3	US-08-510-646B-17
8	32.6	3.6	46899	1	US-08-471-119A-1
9	32.6	3.6	485	2	US-08-403-852D-15
10	32.2	3.6	485	3	US-08-510-646B-15
11	32.2	3.6	4603	1	US-08-258-261B-17
12	32.2	3.6	4603	1	US-08-456-837-17
13	32.2	3.6	4603	1	US-08-457-342-17
14	32.2	3.6	4603	1	US-08-457-646A-17
15	32.2	3.6	4603	1	US-08-458-076A-17
16	32.2	3.6	4603	1	US-08-457-335A-17
17	32.2	3.6	4603	1	US-08-729-214-17
18	32.2	3.6	4603	1	US-09-028-934-17
19	32.2	3.6	5698	1	US-08-761-258-11
20	32.2	3.6	5698	1	US-08-977-306-11
21	32	3.5	7218	1	US-08-232-463-14
22	31.8	3.5	1615	4	US-08-176-413-4
23	31.8	3.5	1615	4	PCT-US94-14919-4
24	31.8	3.5	3396	3	US-08-974-549A-638
25	31.2	3.5	36519	3	US-08-923-137-2
26	31.2	3.5	80161	3	US-09-036-987A-1
27	31	3.4	15664	1	US-08-402-282-3
28	31	3.4	15664	1	US-08-508-004-3

29	31	3.4	15664	1	US-08-402-066-3	Sequence 3, Appli
30	31	3.4	15664	1	US-08-402-068-3	Sequence 3, Appli
31	30.8	3.4	9515	1	US-08-920-812-13	Sequence 13, Appli
32	30.8	3.4	9515	1	US-08-920-827-13	Sequence 13, Appli
33	30.8	3.4	9515	1	US-08-921-177-13	Sequence 13, Appli
34	30.8	3.4	9515	1	US-08-362-577C-13	Sequence 13, Appli
35	30.8	3.4	9515	2	US-08-125-468-1	Sequence 13, Appli
36	30.8	3.4	30001	2	US-08-474-933-1	Sequence 1, Appli
37	30.8	3.4	30001	2	US-08-533-669A-9	Sequence 1, Appli
38	30.6	3.4	1618	2	US-08-607-509-1	Sequence 9, Appli
39	30.6	3.4	1618	2	US-08-454-036-1	Sequence 1, Appli
40	30.6	3.4	1618	2	US-08-634-642-1	Sequence 1, Appli
41	30.6	3.4	1618	3	US-08-989-370-1	Sequence 1, Appli
42	30.6	3.4	1618	4	PCT-US95-05064-1	Sequence 1, Appli
43	30.6	3.4	68750	3	US-09-335-409-1	Sequence 1, Appli
44	30.6	3.4	291	2	US-08-403-852D-16	Sequence 16, Appli
45	30	3.3				

ALIGNMENTS

```

RESULT 1
US-08-125-468-1
Sequence 1, Application US/08125468
Patent No. 5589385
GENERAL INFORMATION:
APPLICANT: Ryan, Michael J.
APPLICANT: Lotvin, Jason A.
APPLICANT: Strachy, Nancy
APPLICANT: Fantini, Susan E.
TITLE OF INVENTION: Cloning of the biosynthetic pathway for
TITLE OF INVENTION: chlorotetracycline and tetracycline formation and cosmid
NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSER: American Cyanamid Company
STREET: One Cyanamid Plaza
CITY: Wayne
STATE: New Jersey
COUNTRY: USA
ZIP: 07470
COMPUTER READABLE FORM:
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/125,468
FILING DATE: 22-SEP-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Tsevdos, Estelle J
REGISTRATION NUMBER: 31,145
REFERENCE/DOCKET NUMBER: 31,255-02
TELECOMMUNICATION INFORMATION:
TELEPHONE: (201)831-3241
TELEFAX: (201)831-3305
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 30001 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-125-468-1

```

Query Match 4.4% Score 40: DB 1: Length 30001:
Best Local Similarity 51.7%: Pred. No. 0.023:
Matches 91: Conservatve 0: Mismatches 85: Indels 0: Gaps 0:
360 gtcgcatcggagcgtccgcatggtatccctcaagaagagcgatccatccatccgcgc 419

Db 9524 GCCCGTCTGAGACGCGCTGAGGCGCACCGCGCTCCGCGCGGCGAGGTGTGAGAGACCGC 9583
QY 420 gatttgacatacaccctcggtccaccgcgaacccgcgcgtgagcggtgagttgacatca 479
Db 9584 CTGCTGCTGAGTTCACGACGCGGACACCGGCGCGCCGAGACGCTCGGCTCAGCCACCG 9643
QY 480 gaacgttcgggttaatttcgaacagctgagtctgacttttgcgataccgacg 535
Db 9644 CAACCTGCTGTGTAACGCGGCGAGTGTGCGGACGCGCCACCGCTGACGCGGCGACG 9699

RESULT 2

US-08-474-933-1
; Sequence 1, Application US/08474933
; Patent No. 5866410
; GENERAL INFORMATION:
; APPLICANT: Ryan, Michael J.
; APPLICANT: Lotvin, Jason A.
; APPLICANT: Strahly, Nancy
; APPLICANT: Fantini, Susan E.
; TITLE OF INVENTION: Cloning of the biosynthetic pathway for
; TITLE OF INVENTION: chlorotetracycline and tetracycline formation and cosmid
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: American Cyanamid Company
; STREET: One Cyanamid Plaza
; CITY: Wayne
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07470
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/474,933
; FILING DATE:
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/125,468
; FILING DATE: 22-SEP-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Tsevdos, Estelle J
; REGISTRATION NUMBER: 31,145
; REFERENCE/DOCKET NUMBER: 31,255-02
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (201)831-3241
; TELEFAX: (201)831-3305
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 30001 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-474-933-1

Query Match 4.4%; Score 40; DB 2; Length 30001;
Best Local Similarity 51.7%; Pred. No. 0.023;
Matches 91; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

QY 360 gctcgatctggagctccgaatggtatattcattcaagaagacgagatattcattcaccg 419
Db 9524 GCCCGTCTGAGACGCGCTGAGGCGCACCGCGCTCCGCGCGGCGAGGTGTGAGAGACCGC 9583
QY 420 gatttgacatacaccctcggtccaccgcgaacccgcgcgtgagcggtgagttgacatca 479
Db 9584 CTGCTGCTGAGTTCACGACGCGGACACCGGCGCGCCGAGACGCTCGGCTCAGCCACCG 9643

QY 480 gaacgttcgggttaatttcgaacagctgagtctgacttttgcgataccgacg 535
Db 9644 CAACCTGCTGTGTAACGCGGCGAGTGTGCGGACGCGCCACCGCTGACGCGGCGACG 9699

RESULT 3

US-09-354-129-7/c
; Sequence 7, Application US/09354129
; Patent No. 6136588
; GENERAL INFORMATION:
; APPLICANT: YE, RICK W.
; TITLE OF INVENTION: GENES ENCODING DENITRIFICATION REACTIONS
; FILE REFERENCE: CL-1255
; CURRENT APPLICATION NUMBER: US/09/354,129
; CURRENT FILING DATE: 1999-07-15
; EARLIER APPLICATION NUMBER: 60/093,181
; EARLIER FILING DATE: July 17, 1998
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 7
; LENGTH: 1899
; TYPE: DNA
; ORGANISM: Pseudomonas sp. strain G-179
US-09-354-129-7

Query Match 4.1%; Score 37; DB 3; Length 1899;
Best Local Similarity 51.5%; Pred. No. 0.058;
Matches 85; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 139 ctcgagtaagctgcgtcgcctatctcgcgcttgcgaagcgcgagcgcgcgttcgcgt 198
Db 927 CTCGTCAGTTCGTGCGGCGCGCGCGCGCGCTGATGATGCTGCTGCGCGCGGCGGTG 868
QY 199 tcggttcacaagcgcgcttaccgataagacgttcgcgttcgactgagtgatctcg 258
Db 867 GACATTAACCATCTCCGCGCATGCGAGGATTTTTCGACCGGTTGAGAAVGAACGGCT 808
QY 259 ccggtgacatctcactacatcgtctgcgttgacgacgctg 303
Db 807 GCGCTCGCTTCTTCTTGAAGATCTTTTCGCGCACGCGCAATGTG 763

RESULT 4

US-08-222-617A-3
; Sequence 3, Application US/08222617A
; Patent No. 5882879
; GENERAL INFORMATION:
; APPLICANT: Veenstra, Annemarie E.
; APPLICANT: Martin, Juan F.
; APPLICANT: Garcia, Bruno D.
; APPLICANT: Gutierrez, Santiago
; APPLICANT: Barredo, Jose L.
; APPLICANT: Von Doehren, Hans
; APPLICANT: Pallisa, Harriet
; APPLICANT: Van Liempt, Henk
; APPLICANT: Montenegro, Eduardo P.
; TITLE OF INVENTION: A method for influencing Beta-lactam
; TITLE OF INVENTION: Antibiotic Production and for Isolation of Large
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
; STREET: 300 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25

TELEPHONE: (202) 408-4000

; PRIOR APPLICATION DATA:

```

? APPLICATION NUMBER: FR 92/11441
? FILING DATE: 25-SEP-1992
? ATTORNEY/AGENT INFORMATION:
? NAME: Meyers, Kenneth J.
? REGISTRATION NUMBER: 25, 146
? REFERENCE/DOCKET NUMBER: 03806, 0054-01000
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (202) 408-4000
? TELEFAX: (202) 408-4400
? INFORMATION FOR SEQ ID NO: 15:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 485 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: double
? TOPOLOGY: linear
? MOLECULE TYPE: CDNA
? HYPOTHETICAL: NO
? ANTI-SENSE: NO
? ORIGINAL SOURCE:
? ORGANISM: S.pristinaespiralis
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 3..485
? OTHER INFORMATION: /product= "Partie du gene Smbp
? OS-08-510-6466-15

```

Query Match	3.6%	Score	32.2	DB	3	Length	485
Best Local Similarity	70.5%	Pred. No.	0.95				
Matches	43	Conservative	0	Mismatches	18	Indels	0
						Gaps	0

Qy 416 ccggtatttgcgaatacaacccctccgggtccacccgcgaacggccgtgcgttgatgtatgcc 475

Db 226 CGGCGTACGTATCCACACACCTCCGGCTCCACCGGCACCCCCAAGGCCGTGCTCATGCCCC 285

QY	476	a	476
		—	
Db	286	A	286

RESULT 11
US-08-258-261B-17

; Sequence 17, Application US/08258261B
; Patent No. 5639949

```

; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas

```

APPLICANT: Ligon, James M.
APPLICANT: Beck, James Joseph

APPLICANT: Hill, Dwight Steven
APPLICANT: Ryals, John Andrew

APPLICANT: Gaffney, Thomas Deane
APPLICANT: Lam, Stephen Ting

APPLICANT: Hammer, Phillip E.
APPLICANT: Uknes, Scott Joseph

TITLE OF INVENTION: Genes for the
TITLE OF INVENTION: antipathogeni

NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:

ADDRESSEE: Ciba-Gelgy Corporati
STREET: 7 Skyline Drive

CITY: Hawthorne
STATE: NY

STATE: NI
COUNTRY: USA
STR: 10533

```

; ZIF: 10532
;
; COMPUTER READABLE FORM:
;

```

```

; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible

```

```
;; OPERATING SYSTEM:  PC-DOS/MS-DOS
;; SOFTWARE:  PatentIn Release #1.0
```

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/258,2

```

;
; FILING DATE: 08-JUN-1994
; CLASSIFICATION: 800
;

1 PRIOR APPLICATION DATA:
 2 APPLICATION NUMBER: US 08/457,205
 3 FILING DATE: 01-JUN-1995
 4 ATTORNEY/AGENT INFORMATION:
 5 NAME: Elmer, James SCOTL
 6 REGISTRATION NUMBER: 36,129
 7 REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
 8 TELECOMMUNICATION INFORMATION:
 9 TELEPHONE: 919-541-8614
 10 TELEFAX: 919-541-8689
 11 INFORMATION FOR SEQ ID NO: 17:
 12 SEQUENCE CHARACTERISTICS:
 13 LENGTH: 4603 base pairs
 14 TYPE: nucleic acid
 15 STRANDEDNESS:
 16 TOPOLOGY: linear
 17 MOLECULE TYPE: DNA (genomic)
 18 FEATURE:
 19

```
NAME/KEY: CDS
LOCATION: 230..1594
OTHER INFORMATION: /gene="phz1"
OTHER INFORMATION: /label=ORF1
```

```

? OTHER INFORMATION: /note= "open Reading Frame #1 for DNA sequence
FEATURE:
? NAME/KEY: CDS
? LOCATION: 1598..2758

```

```

OTHER INFORMATION: /gene= "p1u2"
OTHER INFORMATION: /label= ORF2
OTHER INFORMATION: /note= "Open Reading Frame # 2 for DNA sequence"
FEATURE:

```

```

: NAME/KEY: CDS
: LOCATION: 276..3597
: OTHER INFORMATION: /gene= "phz3"
: OTHER INFORMATION: /label= ORF3

```

```

; OTHER INFORMATION: /note="Open Reading Frame #3 for DNA sequence"
;
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 3597..4262
;

```

```

OTHER INFORMATION: /label= ORF4
OTHER INFORMATION: /note= "Open Reading Frame #4 of DNA sequence.
OTHER INFORMATION: is repeated in SEQ ID NO:21 due to overlapping
FEATURE:

```

```

NAME/KEY: misc_feature
LOCATION: 1..4603
OTHER INFORMATION: /note="Four open reading frames
OTHER INFORMATION:

```

OTHER INFORMATION: Example 18 of the specification."
US-08-258-261B-17

Query Match	3.6%	Score 32.2;	DB 1;	Length 4603;
Best Local Similarity	49.7%;	Pred. No. 2.6;		
Matches 82; Conservative	0;	Mismatches 83;	Indels 0;	Gaps

Dy 114 cgttgtagctctctcctccgcaggaactcgagtacgctcgccatctcggcgcttqca 173
| | | | | | | | | | | | | | | | | | |
Db 961 CGGCGAATGTCTGGACCCGAGGTCTTGCCTGCCGATTGACTTGCTGCCACTTCGATCGG 1020

Dy 174 ggcgggagcgcacgcgcgttgcgcgcttcggtaccacaaggcggtaccgaatgaagttc 233
Db 1021 GCGGCCTGCGCTGGATGTACTGGCGCTGATCCCTTACCGCAGATCGCGCAACGCCGTTT 108

OY 234 cgatcgcgtactgtagtgccttcgcgcgtggccattcctacac 278
||| ||| ||| ||| ||| ||| ||| |||
Db 1081 CGAGCGCGTGGACGATGATGTCGCCGTCTGGCATGAAATCACC 1125

RESULT 12
US-08-456-837-17

```

; sequence 1//, application US/0842083/
; Patent No. 5643774
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
;

```

```

OTHER INFORMATION: /note= "Four open reading frames
OTHER INFORMATION:
OTHER INFORMATION: Example 18 of the specification."
US-08-456-837-17

Query Match          3.6%: Score 32.2; DB 1; Length 4603;
Best Local Similarity 49.7%; Pred. No. 2.6;
Matches 82; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

QY 114 cgtgtgatccctctcccgacgaggaactcgatagctgcgcctatctccgcgcttgcga 173
      || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 961 CGGGAATGTGCGGAACCGCAGGTCTCGCGGATATGACTTTCCTGCGCACCCTGATCGG 1020

QY 174 ggcgcgggcgcacgcgcgcttgcgccttgcgttccacaaggcgcgcttaccgatlgaagctc 233
      || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1021 CGCGCTCGCTCGATGTACTGCGCGCTGATCTCCCTACCGCAAGATCGCGGAACGGGTTT 1080

QY 234 cgatcgctactgagtgatgtctgtgcgcggttgcacattctactac 278
      || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1081 CGAGCGGTGGACGATGATGCGCGCTGCTGCGATGAACATCAC 1125

RESULT 13
US-08-457-342-17
; Sequence 17, Application US/08457342
; Patent No. 5662898
; GENERAL INFORMATION:
; APPLICANT: Schnupp, Thomas
; APPLICANT: Ligon, James M.
; APPLICANT: Beck, James Joseph
; APPLICANT: Hill, Dwight Steven
; APPLICANT: Ryals, John Andrew
; APPLICANT: Gaffney, Thomas Deane
; APPLICANT: Lam, Stephen Ting
; APPLICANT: Hammer, Phillip E.
; APPLICANT: Uknes, Scott Joseph
; TITLE OF INVENTION: Genes for the synthesis of
; TITLE OF INVENTION: antipathogenic substances
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/457,342
; FILING DATE: 01-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/457,205
; FILING DATE: 01-JUN-1995
; APPLICATION NUMBER: 08/258,261
; FILING DATE: 08-Jun-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8614
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4603 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single

```

```

? TOPOLOGY: linear
? MOLECULE TYPE: DNA (genomic)
? FEATURE:
?   NAME/KEY: CDS
?   LOCATION: 230..1594
?   OTHER INFORMATION: /gene= "phz1"
?   OTHER INFORMATION: /label= ORF1
?   OTHER INFORMATION: /note= "Open Reading Frame #1 for DNA sequence"
?   FEATURE:
?     NAME/KEY: CDS
?     LOCATION: 1598..2758
?     OTHER INFORMATION: /gene= "phz2"
?     OTHER INFORMATION: /label= ORF2
?     OTHER INFORMATION: /note= "Open Reading Frame #2 for DNA sequence"
?     FEATURE:
?       NAME/KEY: CDS
?       LOCATION: 2764..3597
?       OTHER INFORMATION: /gene= "phz3"
?       OTHER INFORMATION: /label= ORF3
?       OTHER INFORMATION: /note= "Open Reading Frame #3 for DNA sequence"
?       FEATURE:
?         NAME/KEY: misc_feature
?         LOCATION: 3597..4262
?         OTHER INFORMATION: /label= ORF4
?         OTHER INFORMATION: /note= "Open Reading Frame #4 of DNA sequence. This informat
?         OTHER INFORMATION: is repeated in SEQ ID NO:21 due to overlapping ORFs."
?         FEATURE:
?           NAME/KEY: misc_feature
?           LOCATION: 1..4603
?           OTHER INFORMATION: /note= "Four open reading frames
?           OTHER INFORMATION:
?           OTHER INFORMATION: Example 18 of the specification."
US-08-457-342-17
```

Query Match 3.6%; Score 32.2; DB 1; Length 4603;
Best Local Similarity 49.7%; Pred. No. 2.6;
Matches 82; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

```

QY 114 cgtggtatctgtcgtccgaggaactcgtgctgcctatctcgcggttga 173
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 961 CGCGGAATGTGGAAACCGAGGTCTGGCCGATATGACTTCCCTCCACCTGATCGG 1020

QY 174 ggcgcggcgatcgccgtcgtcgttcgtcccaagcgcggtaccgatacgttc 233
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1021 CGGCGCTCGCTGATATGATGGCGCTGATCCCTACCGCAAGTCCCGAAGCGGTTT 1080

    234 cgaatcgtaactgagatcgtcgtccggtgacctatctcaactac 278
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    1081 CGAGCGCGTGGACGATGATCGCCGCTGTCGCGATGACATCACC 1125
```

```

RESULT 14
US-08-457-646A-17
; Sequence 17, Application US/08457646A
; Patent No. 5679560
; GENERAL INFORMATION:
;   APPLICANT: Schupp, Thomas
;   APPLICANT: Ligon, James M.
;   APPLICANT: Beck, James Joseph
;   APPLICANT: Hill, Dwight Steven
;   APPLICANT: Ryals, John Andrew
;   APPLICANT: Gaffney, Thomas Deane
;   APPLICANT: Lam, Stephen Ting
;   APPLICANT: Hammer, Phillip E.
;   TITLE OF INVENTION: Genes for the synthesis of
;   TITLE OF INVENTION: antipathogenic substances
;   NUMBER OF SEQUENCES: 22
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Ciba-Geigy Corporation
;   STREET: 7 Skyline Drive
;   CITY: Hawthorne
```

```

? STATE: NY
? COUNTRY: USA
? ZIP: 10532
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patentin Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
?   APPLICATION NUMBER: US/08/457,646A
?   FILING DATE: 01-JUN-1995
?   CLASSIFICATION: 530
?   PRIOR APPLICATION DATA:
?   APPLICATION NUMBER: US 08/457,205
?   FILING DATE: 01-JUN-1995
?   APPLICATION NUMBER: 08/258,261
?   FILING DATE: 08-Jun-1994
?   ATTORNEY/AGENT INFORMATION:
?   NAME: Elmer, James Scott
?   REGISTRATION NUMBER: 36,129
?   REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
?   TELECOMMUNICATION INFORMATION:
?   TELEPHONE: 919-541-8614
?   TELEFAX: 919-541-8689
?   INFORMATION FOR SEQ ID NO: 17:
?   SEQUENCE CHARACTERISTICS:
?     LENGTH: 4603 base pairs
?     TYPE: nucleic acid
?     STRANDEDNESS: single
?     TOPOLOGY: linear
?     MOLECULE TYPE: DNA (genomic)
?     FEATURE:
?       NAME/KEY: CDS
?       LOCATION: 230..1594
?       OTHER INFORMATION: /gene= "phz1"
?       OTHER INFORMATION: /label= ORF1
?       OTHER INFORMATION: /note= "Open Reading Frame #1 for DNA sequence"
?       FEATURE:
?         NAME/KEY: CDS
?         LOCATION: 1598..2758
?         OTHER INFORMATION: /gene= "phz2"
?         OTHER INFORMATION: /label= ORF2
?         OTHER INFORMATION: /note= "Open Reading Frame #2 for DNA sequence"
?         FEATURE:
?           NAME/KEY: CDS
?           LOCATION: 2764..3597
?           OTHER INFORMATION: /gene= "phz3"
?           OTHER INFORMATION: /label= ORF3
?           OTHER INFORMATION: /note= "Open Reading Frame #3 for DNA sequence"
?           FEATURE:
?             NAME/KEY: misc_feature
?             LOCATION: 3597..4262
?             OTHER INFORMATION: /label= ORF4
?             OTHER INFORMATION: /note= "Open Reading Frame #4 of DNA sequence. This infor
?             OTHER INFORMATION: is repeated in SEQ ID NO:21 due to overlapping ORFs."
?             FEATURE:
?               NAME/KEY: misc_feature
?               LOCATION: 1..4603
?               OTHER INFORMATION: /note= "Four open reading frames
?               OTHER INFORMATION:
?               OTHER INFORMATION: Example 18 of the specification."
US-08-457-646A-17
```

Query Match 3.6%; Score 32.2; DB 1; Length 4603;
Best Local Similarity 49.7%; Pred. No. 2.6;
Matches 82; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

```

QY 114 cgtggtatctgtcgtccgaggaactcgtgctgcctatctcgcggttga 173
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 961 CGCGGAATGTGGAAACCGAGGTCTGGCCGATATGACTTCCCTCCACCTGATCGG 1020

QY 174 ggcgcggcgatcgccgtcgtcgttcgtcccaagcgcggtaccgatacgttc 233
```

Db 1021 CGCGGCTCGCCGTGAGTACTGGCGCTGATCCCTACCGCCGAGATCGCCGAAACGGCGTTT 1080

Qy 234 cgattcggtactgattgctgcgcgcggtggtgacattctactac 278

Db 1081 CGAGCGGTGACGATGATGCTCGCGCTGCTGCGCATGAACATCAC 1125

RESULT 15

US-08-458-076A-17

; Sequence 17, Application US/08458076A

; Patent No. 5698425

; GENERAL INFORMATION:

; APPLICANT: Schupp, Thomas

; APPLICANT: Ligon, James M.

; APPLICANT: Beck, James Joseph

; APPLICANT: Hill, Dwight Steven

; APPLICANT: Ryals, John Andrew

; APPLICANT: Gaffney, Thomas Deane

; APPLICANT: Lam, Stephen Ting

; APPLICANT: Hammer, Phillip E.

; APPLICANT: Uknes, Scott Joseph

; TITLE OF INVENTION: Genes for the synthesis of

; TITLE OF INVENTION: antipathogenic substances

; NUMBER OF SEQUENCES: 22

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: CIDA-Geigy Corporation

; STREET: 7 Skyline Drive

; CITY: Hawthorne

; STATE: NY

; COUNTRY: USA

; ZIP: 10532

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/458,076A

; FILING DATE: 01-JUN-1995

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/457,205

; FILING DATE: 01-JUN-1995

; APPLICATION NUMBER: 08/258,261

; FILING DATE: 08-Jun-1994

; ATTORNEY/AGENT INFORMATION:

; NAME: Elmer, James Scott

; REGISTRATION NUMBER: 36,129

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 919-541-8614

; TELEFAX: 919-541-8689

; INFORMATION FOR SEQ ID NO: 17:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 4603 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 230..1594

; OTHER INFORMATION: /gene= "phz1"

; OTHER INFORMATION: /label= ORF1

; OTHER INFORMATION: /note= "Open Reading Frame #1 for DNA sequence"

; NAME/KEY: CDS

; LOCATION: 1598..2758

; OTHER INFORMATION: /gene= "phz2"

; OTHER INFORMATION: /label= ORF2

; OTHER INFORMATION: /note= "Open Reading Frame #2 for DNA sequence"

; NAME/KEY: CDS

; LOCATION: 2764..3597

; OTHER INFORMATION: /gene= "phz3"

; OTHER INFORMATION: /label= ORF3

; OTHER INFORMATION: /note= "Open Reading Frame #3 for DNA sequence"

; FEATURE:

; NAME/KEY: misc_feature

; LOCATION: 3597..4262

; OTHER INFORMATION: /label= ORF4

; OTHER INFORMATION: /note= "Open Reading Frame #4 of DNA sequence. This infor

; OTHER INFORMATION: is repeated in SEQ ID NO:21 due to overlapping ORFs."

; FEATURE:

; NAME/KEY: misc_feature

; LOCATION: 1..4603

; OTHER INFORMATION: /note= "Four open reading frames

; OTHER INFORMATION:

; OTHER INFORMATION: Example 18 of the specification."

; US-08-458-076A-17

Query Match

Best Local Similarity 3.6%; Score 32.2; DB 1; Length 4603;

Matches 82; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

Qy 114 cgtgtgattctctgtctccgcagggagctcgagtacgtctgctctctcgcgcgttgca 173

Db 961 CGCGGAAATGTCGGAACCGCAGTCTCGCGCATATGACATTGCTCGCTGACCTCGATCGG 1020

Qy 174 ggcgcgcgcacatcgccgcgtgcgccttcggttcacaaagcgcggttacgattgaacgttc 233

Db 1021 CGCGGCTCGCCGTGAGTACTGGCGCTGATCCCTACCGCCGAGATCGCCGAAACGGCGTTT 1080

Qy 234 cgattcggtactgattgctgcgcgcggtggtgacattctactac 278

Db 1081 CGAGCGGTGACGATGATGCTCGCGCTGCTGCGCATGAACATCAC 1125

Search completed: January 12, 2001, 20:14:13

Job time: 26521 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd

```
Run on:      January 12, 2001, 20:05:55 ; Search time 244.98 Seconds
              (without alignments)
              1384.698 Million cell updates/sec
```

Title:	US-09-461-774-5
Perfect score:	903
Sequence:	1 gattacgacagcactgga.....aggagaggtgaaagcttaa 903

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 480022 segs, 187831343 residues
total number of hits satisfying chosen parameters: 960044

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

Post-processing:	Minimum Match	0%
	Maximum Match	100%
	Listing first	45 summaries

Database : N_Geneseq_36: *

1	/SID6/gcgdata/geneseq/geneseqn/NA1960.DAT *
2	/SID6/gcgdata/geneseq/geneseqn/NA1961.DAT *
3	/SID6/gcgdata/geneseq/geneseqn/NA1962.DAT *
4	/SID6/gcgdata/geneseq/geneseqn/NA1963.DAT *
5	/SID6/gcgdata/geneseq/geneseqn/NA1964.DAT *
6	/SID6/gcgdata/geneseq/geneseqn/NA1965.DAT *
7	/SID6/gcgdata/geneseq/geneseqn/NA1966.DAT *
8	/SID6/gcgdata/geneseq/geneseqn/NA1967.DAT *
9	/SID6/gcgdata/geneseq/geneseqn/NA1968.DAT *
10	/SID6/gcgdata/geneseq/geneseqn/NA1969.DAT *
11	/SID6/gcgdata/geneseq/geneseqn/NA1970.DAT *
12	/SID6/gcgdata/geneseq/geneseqn/NA1971.DAT *
13	/SID6/gcgdata/geneseq/geneseqn/NA1972.DAT *
14	/SID6/gcgdata/geneseq/geneseqn/NA1973.DAT *
15	/SID6/gcgdata/geneseq/geneseqn/NA1974.DAT *
16	/SID6/gcgdata/geneseq/geneseqn/NA1975.DAT *
17	/SID6/gcgdata/geneseq/geneseqn/NA1976.DAT *
18	/SID6/gcgdata/geneseq/geneseqn/NA1977.DAT *
19	/SID6/gcgdata/geneseq/geneseqn/NA1978.DAT *
20	/SID6/gcgdata/geneseq/geneseqn/NA1979.DAT *
21	/SID6/gcgdata/geneseq/geneseqn/NA1980.DAT *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	40	4.4	30001	18	T61016	Total DNA sequence
2	40	4.4	30001	20	X05110	S. aureofaciens DN
3	36.2	4.0	1318	19	V49982	Floral organ-speci
4	36.2	4.0	2636	19	V49983	Floral organ-speci
5	35.4	3.9	11601	12	Q13608	ACV synthetase gene
6	35.4	3.9	13058	14	Q48231	Vector containing
7	35	3.9	7673	19	V58229	Omega-cyclohexane
8	34.8	3.9	12441	21	V87264	S. venezuelae desce
9	34.8	3.9	13613	21	287319	S. venezuelae desce
10	34.8	3.9	23666	12	Q10190	Cephalosporin antil
11	34.8	3.9	38506	21	Z56001	Recombinant cosmid
12	34.2	3.8	37856	21	A11992	S. cellulosum DNA

C	13	34	3.8	2000	13	025783	CYP synthase gene
C	14	34	3.6	3001	12	014668	CYP synthase gene
C	15	32.8	3.6	1137	21	A02477	Human colon cancer
C	16	32.8	3.6	1966	20	X04340	Human secreted protein
C	17	32.6	3.6	46899	924	A12829	<i>Pseudomonas</i> <i>cypA</i>
C	18	32.6	3.6	46899	15	054386	<i>T. niyeum</i> CycloSP
C	19	32.4	3.6	13144	12	013288	<i>P. denitrificans</i> g
C	20	32.2	3.6	4602	17	T06770	<i>Pseudomonas aureo</i>
C	21	32.2	3.6	4603	18	T89957	<i>Pseudomonas aureo</i>
C	22	32.2	3.6	4603	19	V56732	Phenazine gene cl
C	23	32.2	3.6	5698	19	V38844	<i>Pseudomonas fluores</i>
C	24	32.2	3.6	5698	20	X93375	<i>P. aeruginosa</i> fluo
C	25	32	3.5	1008	11	006843	Secondary metabol
C	26	32	3.5	1209	21	287293	<i>S. venezuelae</i> des
C	27	32	3.5	2417	15	056106	DNA encoding a li
C	28	32	3.5	8753	12	013285	<i>P. denitrificans</i> E
C	29	31.8	3.5	1413	20	X34205	<i>Mycobacterium spe</i>
C	30	31.8	3.5	1615	16	Q39201	<i>Mycobacterium spe</i>
C	31	31.8	3.5	1722	20	X94206	<i>Mycobacterium spe</i>
C	32	31.6	3.5	2004	14	051113	Lignin peroxidase
C	33	31.2	3.5	36519	19	V22141	Chimpanzee adenov
C	34	31.2	3.5	80161	20	221501	DNA fragment of S
C	35	31.2	3.5	534720	19	V30458	Rhizobium species
C	36	31.2	3.5	536165	17	T03459	Rhizobium species
C	37	31	3.4	15664	17	T09312	<i>Mycobacteriophage</i>
C	38	31	3.4	15664	18	T70492	<i>Mycobacteriophage</i>
C	39	31	3.4	15664	18	T66130	<i>Mycobacteriophage</i>
C	40	31	3.4	15664	18	T51224	NheI-c/SpeI fragm
C	41	30.8	3.4	3635	14	041652	Phosphoglycerate
C	42	30.8	3.4	9515	15	055145	<i>Pseudomonas aerug</i>
C	43	30.8	3.4	30001	18	T61016	Total DNA sequenc
C	44	30.8	3.4	30001	20	X05110	<i>S. aureofaciens</i> D
C	45	30.6	3.4	1145	16	T04785	Cellobiose gene D

ALIGNMENTS

RESULT	1
ID	T61016
	T61016 standard; DNA; 30001 bp.
XX	
AC	T61016;
XX	
DT	21-APR-1997 (first entry)
XX	
DE	Total DNA sequence from cosmid clones LP(2)127 and LP(2)128
XX	
KW	Cosmid clone: LP(2)127; LP(2)128; chlorotetracycline;
KW	biosynthetic pathway; recombinant; production; antibiotic;
KW	heterologous host; Streptomyces lividans; ss.
XX	
OS	Streptomyces aureofaciens.
XX	
PN	US5589385-A.
XX	
PD	31-DEC-1996.
XX	
PF	26-JUL-1990; 90US-0558039.
XX	
PR	22-SEP-1993; 93US-0125468.
PR	26-JUL-1990; 90US-0558039.
PR	26-JUL-1990; 90US-0558040.
PR	15-JAN-1992; 92US-0821109.
PR	15-JAN-1992; 92US-0821419.
XX	
PA	(AMCY) AMERICAN CYANAMID CO.
PI	Fantini SE, Lovcin JA, Ryan MJ, Strathy N;
DR	WPI: 1997-076853/07.
XX	
XT	DNA encoding tetracycline biosynthetic pathway proteins -


```
FT FT /*tag= e
FT FT /label= ORF5
FT FT /function= "acyl-CoA dehydrogenase"
FT CDS 5812..6837
FT FT /*tag= f
FT FT /label= ORF6
FT FT /function= "alcohol dehydrogenase"
PN JP10234376-A.
XX
XX 08-SEP-1998.
PD
XX 28-FEB-1997; 97JP-0046570.
XX
XX 28-FEB-1997; 97JP-0046570.
PR (KIRI ) KIRIN BEVERAGE KK.
XX
XX WPI: 1998-535030/46.
DR P-PSDB: W71633, W71634, W71635, W71636, W71637, W71638.
XX
XX New nucleic acid - useful for detection and identification of genus
XX Alicyclobacillus microorganism(s)
XX
XX Claim 7: Page 9-18; 37pp: Japanese.
XX
XX The present sequence represents a new nucleic acid which encodes enzymes
XX which participate in the biosynthesis of Omega-cyclohexane fatty acid.
XX The nucleic acid is isolated from Alicyclobacillus acidocaldarius. The
XX present invention also describes primers and probes containing all or
XX part of the nucleic acid from Alicyclobacillus acidocaldarius. The
XX primers and probes may be used for detection and/or identification of a
XX microorganism of genus Alicyclobacillus. The method can detect and
XX identify Alicyclobacillus genus rapidly and easily.
SQ
XX
XX Sequence 7673 BP; 1288 A; 2432 C; 2762 G; 1191 T; 0 other;
XX
XX Query Match 3.9%; Score 35; DB 19; Length 7673;
XX Best Local Similarity 56.5%; Pred. No. 1.2;
XX Matches 65; Conservative 0; Mismatches 50; Indels 0; Gaps 0;
XX
XX QY 102 cagcggtgacgcggtgattctctgctccgcagagcagtcgagtcgctcattct 161
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX Db 3065 caaggggagacacgcgtgcgtgcagctgcgaaagtcgctcgaattcgttgccgtgtt 3124
XX
XX QY 162 cggcgagtcgacggcgagcgacatcgcgctgcccgttcggttcacaaaggcgcc 216
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 3125 cgctgcgagcgatcgagcgacatcgcgctgctgcgcgacatcgatctcgccgc 3179
XX
XX RESULT 8
XX 287284 287284 standard; DNA: 12441 BP.
XX
XX 287284;
XX
XX 05-JUN-2000 (first entry)
XX
XX S. venezuelae desosamine biosynthetic gene cluster pikB, SEQ ID NO:3.
XX
XX Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
XX neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
XX biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma,
XX chronic obstructive pulmonary disease; respiratory inflammation;
XX hypercholesterolemia; crop protection agent; ds.
XX
XX Streptomyces venezuelae ATCC15439.
XX
XX W0200000620-A2.
XX
XX 06-JAN-2000.
XX
```

```
PF 25-JUN-1999; 99WO-US14398.
XX
XX 26-JUN-1998; 98US-0105537.
PR
XX
XX (MIND ) UNIV MINNESOTA.
XX
XX Sherman DH, Liu H, Xue Y, Zhao L;
XX
XX WPI: 2000-150679/14.
DR P-PSDB: Y77179.
XX
XX Desosamine and macrolide biosynthetic gene clusters, useful for, e.g.
XX synthesis of methymycin and pikromycin -
XX
XX Claim 2: Page 281-287; 438pp; English.
XX
XX The invention relates to an isolated and purified nucleic acid segment
XX comprising a desosamine biosynthetic gene cluster, a fragment or its
XX biologically active variant, where the nucleic acid sequence is not
XX derived from the eryC gene cluster of Saccharopolyspora erythraea or
XX Streptomyces antibiotics. The invention also relates to a macrolide
XX biosynthetic gene cluster, or fragments thereof. The macrolide
XX biosynthetic gene cluster encodes proteins which synthesize methymycin,
XX pikromycin, neomethymycin, narbomycin or a combination of these
XX compounds. Recombinant or augmented cells comprising the desosamine
XX and/or macrolide biosynthetic gene clusters are useful for the production
XX of biologically active macrolides. The macrolide biosynthetic proteins
XX are useful for synthesis of methymycin, pikromycin, neomethymycin and
XX narbomycin. The alternative termination of polyketide synthesis may be
XX useful to prepare novel antibiotics and polyhydroxyalkanoate (PHA)
XX monomers. The compounds produced by the recombinant host cells are useful
XX as biopolymers, e.g., in packaging or biomedical applications, to
XX engineer PHA monomer syntheses or to prepare biologically active agents,
XX such as chemotherapeutics, immunosuppressants, agents to treat asthma,
XX chronic obstructive pulmonary disease as well as other diseases involving
XX respiratory inflammation, cholesterol-lowering agents or macrolide-based
XX antibiotics which are active against a variety of organisms, e.g.,
XX bacteria, including multi-drug resistant pneumococci and other
XX respiratory pathogens, as well as viral parasitic pathogens, or as crop
XX protection agents (e.g., fungicides or insecticides) via expression of
XX polyketide in plants. The present sequence represents the desosamine
XX biosynthetic gene cluster from Streptomyces venezuelae ATCC 15439.
SQ
XX
XX Sequence 12441 BP; 1704 A; 4294 C; 4686 G; 1757 T; 0 other;
XX
XX Query Match 3.9%; Score 34.8; DB 21; Length 12441;
XX Best Local Similarity 50.0%; Pred. No. 1.7;
XX Matches 113; Conservative 0; Mismatches 112; Indels 1; Gaps 1;
XX
XX QY 634 cccggtgtctcacagcccggtgtctctcgcagcgcccgccggtgagatca-ctt 692
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX Db 37 cccgcggcgccgcagatgcgcgcgtctgcagcccccagccattactgaagggagcg 96
XX
XX QY 693 gatgcccagcattctcacgccccttcgcgacgacccaatttcgcttgaactagcgcc 752
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX Db 97 cgaagagatggcagaaccccgcgtgacgcagcccgaggggagccctcaagcagccccc 156
XX
XX QY 753 acgaagaacacccagcagacatggcgcggtgcgtgacatcggaatactgacatct 812
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX Db 157 gctgggctgcgacccctgcgcggttgccgacgctgtaactgacccactcttgtagac 216
XX
XX QY 813 cagcggtagcagcggtgacagcgcgacagatcgaagcgcttcgc 838
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX Db 217 ccgcgcatcactgattcaccgcgcgacgagcgacccgtacgc 262
XX
XX RESULT 9
XX 287319 287319 standard; DNA: 13613 BP.
XX
XX 287319;
XX
```

DT	05-JUN-2000	(first entry)
DE	S. venezuelae	desosamine biosynthetic gene cluster pikB.
KW	Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;	
KM	nemethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;	
KV	biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma,	
KW	chronic obstructive pulmonary disease; respiratory inflammation;	
KX	hypercholesterolaemia; crop protection agent; ds.	
OS	Streptomyces venezuelae ATCC15439.	
PH	Key	Location/Qualifiers
FT	CDS	1..809
FT	/tag= a	/partial
FT	/product=	"pikB gene cluster protein #1 (Y77204)"
FT	/note=	"No initiation codon given in the specification"
FT	806..2014	
FT	//tag= b	
FT	/product=	"pikB gene cluster protein #2 (Y80998)"
FT	2162..6741	
FT	//tag= c	
FT	/product=	"pikB gene cluster protein #3 (Y77205)"
FT	6834..7402	
FT	/tag= d	
FT	/product=	"pikB gene cluster protein #4 (Y77206)"
FT	/transl_except=	(pos:6837..6841, aa:Gln) 7492..8205
FT	//tag= e	
FT	/product=	"pikB gene cluster protein #5 (Y77207)"
FT	complement (7942..8205)	
FT	/partial	
FT	//tag= f	
FT	/product=	"pikB gene cluster protein #6 (Y77208)"
FT	/note=	"No termination codon given in the specification"
FT	/transl_except=	(pos:8270..8272, aa:Val) /transl_except= (pos:8273..8275, aa:Thr) /transl_except= (pos:8276..8278, aa:Gly) complement (10126..11139)
FT	//tag= g	
FT	/product=	"pikB gene cluster protein #7 (Y80999)"
FT	complement (11271..12149)	
FT	//tag= h	
FT	/product=	"pikB gene cluster protein #8 (Y77209)"
FT	complement (12342..13799)	
FT	//tag= i	
FT	/product=	"pikB gene cluster protein #9 (Y77210)"
FT	complement (13706..15043)	
FT	//tag= j	
FT	/product=	"pikB gene cluster protein #10 (Y77211)"
FT	15404..15574	
FT	//tag= k	
FT	/product=	"pikB gene cluster protein #11 (Y77212)"
XX	WO200000620-A2.	
PN	06-JAN-2000.	
XX	25-JUN-1999;	99MO-US14398.
PE	26-JUN-1998;	98US-0105537.
PR	(MINU) UNIV MINNESOTA.	
PA	Sherman DH, Liu H, Xue Y, Zhao L;	
PJ	WP1: 2000-160679/14.	
XX	P-PSDB: Y77204, Y77205, Y77206, Y77207, Y77208, Y77209,	
DR	Y77210, Y77211, Y77212, Y80998, Y80999.	
PT	Desosamine and macrolide biosynthetic gene clusters, useful for, e .g. synthesis of methymycin and pikromycin -	

XX Disclosure: Figure 32; 438pp; English.

PS

CC The invention relates to an isolated and purified nucleic acid segment
CC comprising a desosamine biosynthetic gene cluster, a fragment or its
CC biologically active variant, where the nucleic acid sequence is not
CC derived from the eryC gene cluster of *Saccharopolyspora erythraea* or
CC Streptomyces antibiotics. The invention also relates to a macrolide
CC biosynthetic gene cluster, or fragments thereof. The macrolide
CC biosynthetic gene cluster encodes proteins which synthesise methymycin,
CC pikromycin, neomethymycin, nabomycin or a combination of these
CC compounds. Recombinant or augmented cells comprising the desosamine
CC and/or macrolide biosynthetic gene clusters are useful for the production
CC of biologically active macrolides. The macrolide biosynthetic proteins
CC are useful for synthesis of methymycin, pikromycin, neomethymycin and
CC nabomycin. The alternative termination of polyketide synthesis may be
CC useful to prepare novel antibiotics and polyhydroxyalkanoate (PHA)
CC monomers. The compounds produced by the recombinant host cells are useful
CC as biopolymers, e.g., in packaging or biomedical applications, to
CC engineer PHA monomer synthases or to prepare biologically active agents,
CC such as chemotherapeutics, immunosuppressants, agents to treat asthma,
CC chronic obstructive pulmonary disease as well as other diseases involving
CC respiratory inflammation, cholesterol-lowering agents or macrolide-based
CC antibiotics which are active against a variety of organisms, e.g.,
CC bacteria, including multi-drug resistant pneumococci and other
CC respiratory pathogens, as well as viral parasitic pathogens, or as crop
CC protection agents (e.g., fungicides or insecticides) via expression of
CC polyketides in plants. The present sequence represents the desosamine
CC biosynthetic gene cluster (plik) from *Streptomyces venezuelae* ATCC
CC 15439, as given in figure 32.

XX

SO Sequence 13613 BP; 1858 A; 4733 C; 5092 G; 1930 T; 0 other:

XX

Query Match 3.9%; Score 34.8; DB 21; Length 13613;
Best Local Similarity 50.0%; Pred. No. 1.8;

Matches 113; Conservative 0; Mismatches 112; Indels 1; Gaps 1;

QY 634 ccgcggtgtctaccagcccggtgtcgttcctctcagcgcgccggccggtgatga-clt 692
||||| - | | | | | ||||| - | | | | |
DB 724 ccgcggcgccgcgatgcgcgctcgtcgtacgccccagcacagactltagaaggagcg 783

QY 693 gatgcgcgagatttaacgccttttgcaagcacagaaatcggccttgaactc.tlycgcg 752

DB 794 cgaagaagatlgagaagaccacacggtgacgacactgacggggcccttaacgacgccccc 843

QY 753 acgaagaacaaccgcgcgcagcatgcccggcggtgacctgcgcacaatactgacatcct 812
||| | | | | | | | | ||||| ||||| ||||| | |
DB 844 gctggcgccgacatgcgtccgcggtgtgcccgaacgcgttaactcggcaccacactctctggagac 903

QY 813 cagcggtagcgagcgggtacagcgccgcgaagatacaagcgtctgcgc 858
| | | | | | | | | | | | | | | | | | | |
DB 904 ccgcgcatccactgatarccaacgcgcgaagcgagcccgtagccg 949

RESULT 10
ID Q10190
ID Q10190 standard; DNA; 23666 BP.
XX
AC Q10190:
XX
DT 27-MAR-1991 (first entry)
DE Cephalosporin antibiotic biosynthetic genes.
XX
KW cephalosporin; antibiotic;
KW S-(L)-alpha-aminoadipyl-L-cysteinyI-D-; valine synthetase;
KW isopenicillin N synthetase; isopenicillin N epimerase;
KW deacetoxycephalosporin C synthetase; beta-lactamase;
KW deacetoxycephalosporin C hydroxylase; ss.
XX
SS Lysobacter lactamgenus.

```

FH Key Location/Qualifiers
FT CDS 5524..16692
FT /tag= a
FT /label= ORF 1
FT CDS 16761..17741
FT /tag= b
FT /label= ORF 2
FT CDS 17802..18761
FT /tag= c
FT /label= ORF 3
FT CDS 18798..19739
FT /tag= d
FT /label= ORF 4
FT CDS 19802..21061
FT /tag= e
FT /label= ORF 5
FT CDS 21186..22343
FT /tag= f
FT /label= ORF 6
FT CDS 4211..5443
FT /tag= g
FT /label= ORF 7
FT CDS 2887..4200
FT /tag= h
FT /label= ORF 8
FT CDS 6..2819
FT /tag= i
FT /label= ORF 9

JP02291274-A.
PN
PD 03-DEC-1990.
XX
XX 10-JAN-1990; 90JP-0003762.
XX
XX 01-FEB-1989; 89JP-0024710.
XX 10-JAN-1990; 90JP-0003762.
XX
XX (TAKE ) TAKEDA CHEMICAL IND KK.
XX
XX WPI; 1991-018854/03.
XX DR P-PSDB; R10145,R10688-R10695.
XX
XX Prepnp. of cephalosporin series antibiotics - comprises culturing
XX PT transformant of microbe transformed by plasmid contg. new DNA
XX PT fragment
XX
XX Claim 4; Fig 1; 67pp; Japanese.
XX
XX A fragment of the sequence comprising at least one of the
XX CC cephalosporin biosynthetic enzymes listed in the KEYWORDS can be
XX CC cloned in a plasmid and used to transform microbes, such as
XX CC bacteria or yeast. Although DNA is preferably isolated from L.
XX CC lactamgenus, similar sequences could be obtained from other
XX CC bacteria containing genes coding for biosynthesis of cephalosporin
XX CC series antibiotics. See also Q10191-2.
XX
XX Sequence 23666 BP; 3952 A; 8522 C; 7603 G; 3589 T; 0 other:
SQ

Query Match 3.9%; Score 34.8; DB 12; Length 23666;
Best Local Similarity 53.7%; Pred. No.2.2; Indels 0; Gaps 0;
Matches 72; Conservative 0; Mismatches 62;

OY 417 cgcgtattgcatactccgcgcgttccacccgcgcgcgtgcgtgattgtccca 476
||| ||| | ||||| ||||| ||||| | ||||| |||
DB 13296 cgcgtacgcgtactactccgcgcacacccgcgcgcgcgcgtgcgtgcgcga 13355
||| ||| | ||||| ||||| ||||| | ||||| |||
OY 477 tcagaacgttcgggttaatttcgaacagctgattctgctacttcgcataccgcagcg 536
| ||||| | ||| ||| ||||| | ||| |||
DB 13356 ccgaacagctgcgcgttcgaacgcgcgcgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 13415
||| ||| | ||||| ||||| ||||| | ||||| |||
OY 537 gattccaccgcgcga 550

```

```

DB 13416 ggcgcgcgcgcgcgcga 13429
RESULT 11
ID 256001 standard; DNA; 38506 BP.
XX
XX 256001;
XX
XX 23-MAR-2000 (first entry)
XX
XX Recombinant cosmid pKOS023-27 containing S. venezuelae PKS genes.
XX
XX Narbonolide polyketide synthase; PKS; cosmid pKOS023-27; ketolide;
XX KW antibiotic production; narbomycin; picromycin; ds.
XX
XX Streptomyces venezuelae.
XX
XX Key Location/Qualifiers
XX CDS 70..13725
XX /tag= a
XX /product= PICAI
XX /note= "Narbonolide synthase subunit 1"
XX CDS 13830..25049
XX /tag= b
XX /product= PICAI1
XX /note= "Narbonolide synthase subunit 2"
XX CDS 25133..29821
XX /tag= c
XX /product= PICAI11
XX /note= "Narbonolide synthase subunit 3"
XX CDS 29924..33964
XX /tag= d
XX /product= PICAIIV
XX /note= "Narbonolide synthase subunit 4"
XX CDS 33961..34806
XX /tag= e
XX /product= picB
XX /note= "Contains typeII thioesterase domain"
XX CDS 34863..36011
XX /tag= f
XX /product= PICCII
XX /note= "4-keto-6-deoxyglucose isomerase"
XX CDS 36159..37439
XX /tag= g
XX /product= PICCIII
XX /note= "Desosaminyl transferase"
XX CDS 37529..38242
XX /tag= h
XX /product= PICCVI
XX /note= "3-amino dimethyltransferase"
XX
XX W09961599-A2.
XX
XX 02-DEC-1999.
XX
XX 27-MAY-1999; 99WO-US11814.
XX
XX 28-MAY-1998; 98US-0087080.
XX 28-AUG-1998; 98US-0141908.
XX 22-SEP-1998; 98US-0100880.
XX 08-FEB-1999; 99US-0119139.
XX
XX (KOSA-) KOSAN BIOSCIENCES INC.
XX
XX Ashley G, Betlach MC, Betlach M, McDaniel R, Tang L;
XX WPI; 2000-072618/06.
XX DR P-PSDB; Y67201, Y67202, Y67203, Y67204, Y67205, Y67207, Y67208, Y67211.
XX PT New recombinant DNA encoding a domain of narbonolide polyketide
XX synthase, for production of ketolide antibiotics -

```


Human colon cancer cell line polynucleotide sequence SEQ ID NO: 2468

XX Human: colon cancer; tumour; diagnosis; gene expression product;
KW probe; detection; cancerous state; metastasis; identification;
KW breast cancer; oestrogen receptor-positive breast cancer; therapy;
KW oestrogen receptor-negative breast cancer; lung cancer; ss.

OS Homo sapiens.

XX MO958675-AZ.

XX 18-NOV-1999.

XX 13-MAY-1999; 99WO-US10602.

XX 14-MAY-1998; 98US-0085426.

XX 15-MAY-1998; 98US-0085537.

XX 21-OCT-1998; 98US-0085696.

XX 27-OCT-1998; 98US-0105234.

XX (CHIR) CHIRON CORP.

XX (HYSE-) HYSEQ INC.

PI Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;
PI Reinhard C, Giese K, Randazzo F, Kennedy GC, Pot D, Kassam A;
PI Lamson G, Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I;
PI Leshkowitz D, Kita D, Garcia V, Jones LM, Stache-Crain B;

XX WPI; 2000-126369/11.

XX Polynucleotide library used to determine cancerous states of mammalian
PT cells -

PS Claim 1; Page 990; 1097pp; English.

XX A00010 to A02716 represent polynucleotides isolated from cDNA libraries
CC constructed from human colon cancer cell lines. The present invention
CC also describes a method of detecting differentially expressed genes
CC correlated with a cancerous state of a mammalian cell, comprising
CC detecting at least one differentially expressed gene product in a test
CC sample derived from a cell suspected of being cancerous, where detection
CC of the differentially expressed gene product is correlated with a
CC cancerous state of the cell from which the test sample was derived.
CC The polynucleotides sequences can be used in a method for detecting
CC differentially expressed genes correlated with a cancerous state of a
CC mammalian cell. The polynucleotides can also be used as probes for
CC detecting and mapping related genes. They can be used in diagnosis and
CC prognosis of diseases and disorders (e.g. identification of
CC pre-metastatic or metastatic cancerous states, stages of cancer, or
CC responsiveness of cancer to therapy). This is particularly for breast
CC cancer, oestrogen receptor-positive breast cancer, oestrogen receptor-
CC negative breast cancer, lung cancer, and colon cancer.

XX Sequence 1127 BP; 3 A; 214 C; 505 G; 49 T; 356 other;

XX Query Match 3.6%; Score 32.8; DB 21; Length 1127;

XX Best Local Similarity 38.0%; Pred. No. 2.7; Mismatches 124; Indels 0; Gaps 0;

XX Matches 76; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

QY 17 gggacgagcgtgtgataacccctgacgcgtgacgtgtatcgcgacgacgtgaatgtg 76

DB 745 ggggngcngngngngngcgcgnggnggggagcngcngcgnggngcnggngcng 804

QY 77 cacagagcgtgacgctgtgtgttccacaggtgacgcgtgtgtgacgtgtgtcgcag 136

DB 805 cacngngngngngcnggggngcgcnngngggnggcgtgtgcccncnctngcngcng 864

QY 137 gactcgatcagcgtcgtcgtatctcgtcgtgtgacgagcgagcgagcgtgtgcgc 196

DB 865 cngcngnggggngcngcngcnggngntgtgngggcctggcgngcnncccgngcng 924

QY 197 ttctggttccacaagcgcg 216

DB 925 ncnngncgcgcgcgngngc 944

Search completed: January 12, 2001, 20:07:53
Job time: 26357 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 12, 2001, 20:03:20 ; Search time 6876.06 Seconds
(without alignments)
672.090 Million cell updates/sec

Title: US-09-461-774-5
Perfect score: 903
Sequence: 1 gattacgaacgactgga.....aggagagggtgaacttaa 903

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 118133 seqs, 255875100 residues

Optimal number of hits satisfying chosen parameters: 2236266

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_om:*
4: gb_om:*
5: gb_ov:*
6: gb_pi1:*
7: gb_pi2:*
8: gb_pi3:*
9: gb_pi4:*
10: gb_pi5:*
11: gb_pi6:*
12: gb_pi7:*
13: gb_pi8:*
14: gb_pi9:*
15: gb_pi10:*
16: gb_pi11:*
17: gb_pi12:*
18: gb_pi13:*
19: gb_pi14:*
20: gb_pi15:*
21: gb_pi16:*
22: gb_pi17:*
23: gb_pi18:*
24: gb_pi19:*
25: gb_pi20:*
26: gb_pi21:*
27: gb_pi22:*
28: gb_pi23:*
29: gb_pi24:*
30: gb_pi25:*
31: gb_pi26:*
32: gb_pi27:*
33: gb_pi28:*
34: gb_pi29:*
35: gb_pi30:*
36: gb_pi31:*
37: gb_pi32:*
38: gb_pi33:*
39: gb_pi34:*
40: gb_pi35:*
41: gb_pi36:*
42: gb_pi37:*
43: gb_pi38:*

44: em_hcg2:*
45: em_hcg3:*
46: em_hcg4:*
47: em_hcg5:*
48: em_hcg6:*
49: em_hcg7:*
50: em_hcg8:*
51: em_hcg9:*
52: em_hcg10:*
53: em_hcg11:*
54: em_hcg12:*
55: em_hcg13:*
56: em_hcg14:*
57: em_hcg15:*
58: em_hcg16:*
59: em_hcg17:*
60: em_hcg18:*
61: em_hcg19:*
62: em_hcg20:*
63: em_hcg21:*
64: em_hcg22:*
65: em_hcg23:*
66: em_hcg24:*
67: em_hcg25:*
68: em_hcg26:*
69: em_hcg27:*
70: em_hcg28:*
71: em_hcg29:*
72: em_hcg30:*
73: em_hcg31:*
74: em_hcg32:*
75: em_hcg33:*
76: em_hcg34:*
77: em_hcg35:*
78: em_hcg36:*
79: em_hcg37:*
80: em_hcg38:*
81: em_hcg39:*
82: em_hcg40:*
83: em_hcg41:*
84: em_hcg42:*
85: em_hcg43:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	891.8	98.8	24292	2	MTCY24G1	28358 Mycobacteri
2	888.6	98.4	1749	2	M8U75685	U75685 Mycobacteri
3	870.2	96.4	9699	2	MSGMRKCSYN	M95808 Mycobacteri
4	426.2	47.2	30352	2	MTCY409	Z97188 Mycobacteri
5	426.2	47.2	41321	2	MSGY409	AD000017 Mycobacte
6	400	44.3	11042	2	MTCY045	AL022000 Mycobacte
7	385	42.6	11430	2	MTCY19G5	Z77826 Mycobacteri
8	385	42.6	37316	2	MSGY456	AD000001 Mycobacte
9	371.8	41.2	41171	73	U00010	U00010 Mycobacteri
10	367.4	40.7	37840	2	MTCY005	AL010186 Mycobacte
11	320.2	35.5	43523	2	MTCY349	Z83018 Mycobacteri
12	317.2	35.1	31176	2	MTCY00024	U00024 Mycobacteri
13	317.2	35.1	31176	2	MSGY2	AD000009 Mycobacte
14	313.4	34.7	29372	2	MTCY338	Z74697 Mycobacteri
15	291	32.2	34316	2	MTCB12	AL035480 Mycobacte
16	185	20.5	22570	2	MTCY4D9	Z84725 Mycobacteri
17	185	20.5	42741	2	MSGY423	AD000014 Mycobacte
18	124	13.7	23740	2	MTCY026	AL022076 Mycobacte
19	117.8	13.0	357	2	AF288926S1	AF288926 Mycobacte
20	93.8	10.4	23730	73	SC3F7	AL021409 Streptomy
21	86.6	9.6	22444	1	AE004669	AE004669 Pseudomon


```
misc_feature complement(203..250)
/gene="mas"
/note="PS00012 Phosphopantetheine attachment site"
source 287..23998
/organism="Mycobacterium tuberculosis"
/strain="H37Rv"
/db_xref="taxon:1773"
/clone="Y24G1"
complement(5861..5911)
/gene="mas"
/note="PS00606 Beta-ketoacyl synthases active site"
7032..8774
/gene="fad28"
7032..8774
/gene="fad28"
/note="RV2941, (MTCY24G1.08c), len: 580 aa, fad28,
Probable Acyl-CoA Synthetase, similar to many e.g.
Y06J_MYCTU_Q10976 hypothetical 67.9 kd protein cy338.19
(626 aa), fasta scores, opt: 2337, E(): 0, (57.7% identity
in 619 aa overlap), also MTCY19G5.07, (648 identity in 584
aa overlap), also near identical to YMA2_MYCBO_Q02278
hypothetical 32.3 kd protein in mas 5 (298 aa), fasta
scores, opt: 1884, E(): 0, (99.0% identity in 288 aa
overlap); contains PS00018 EF-handcalcium-binding domain"
/codon_start=1
/transl_table=11
/product="fad28"
/protein_id="CAB06100.1"
/db_xref="GI:1781172"
/db_xref="SPTREMBL:P96290"
/translation="MSVRSIPALRACRIOPHPAPFTFMDYEDMDGVAITITWSQL
YRRTINVAOELSRGSGTGRVYISAPGLEEVYATLGLAQRIAVPLSVQGGTDE
RSDSLSDSPVAILITSSAVDDVQHVARRGESPSITIEDLIDLAPMGTYKED
EYPSAYLQYTSSTRTPAGVYMSHQNVNVEEOLMSGFADTDIIPPNSALWMLP
FYHMDGLVIGICAPILIGYPALVSPVSLQRPAMHMLMSDFHAFSAANFAGELA
ARRTDDMDMAGDGLNILLTSGSERVOATIKRPADEPARNLEOERYIRPSYGLAE
TYYVATSKRQGPPEVDPDTESLSAGHAKPCAGADBARLISMLPRSPITVRVDSDC
IECPGTYGEIHWGHNDYANNGWKPKPDSERFSGKITYPSCGTPEGMWLTGDSGFV
TDKMFITIGRIKDLIVTGRNHSPPDITATTOETTRGCALISVPGDSTELVAIIE
LKKRQSDQDAMARLIGATREVTSALSSHGLSVADLVLAAGSIPITTSKVRGAC
VBOYRQDQFARLDA"
8520..8558
/note="fad28"
/gene="fad28"
/note="PS00018 EF-hand calcium-binding domain"
8767..11529
/gene="mmpL7"
/note="RV2942, (MTCY24G1.07c), len: 920 aa, mmpL1, member
of conserved Mycobacterial Membrane Protein. Similarity to
gpI.A1021932|MTV037_14 Mycobacterium tuberculosis seque
(967aa) opt: 481 z-score: 521.3 E(): 1.6e-21; 25.9%
identity in 922 aa overlap; and to yv34_MYCUE_P54881
hypothetical105.2 kd protein u1740v. (959 aa), fasta
scores, opt: 437, E(): 8.6e-19, (25.8% identity in 923 aa
overlap); contains PS00079 Multicopper oxidases signature
1, PS00044 Bacterial regulatory proteins, lysR family
signature"
/codon_start=1
/transl_table=11
/product="mmpL7"
/protein_id="CAB06107.1"
/db_xref="GI:1781171"
/db_xref="SPTREMBL:P96289"
/translation="MSPDAGRIHRIYIRLKKSPDCRATTITSGSADGORSRLTNL
LVVAAMVAAYIANLILFTQAEPHDTSPALPPOKAAATSRARQAFPGGSAAIA
LYVVGSGTLEPDOPDYDAVSVALRAATRVGSGPVPATRGSTDSGRATA
MYMLRGEGSTOAAESLDAVSVALRQLEPSSIGRASITVPATINDMPQITAMOSATI
VTVAAYIVALLLRARISVRAAATIVLTADISLAVAMPVLAVRGHDGTSVESWTI
AAVLITGITATATMLAARLSDAGSHAAPYRDSIPALPGACVAFITGPDLARTP
ALHGVGTGIVFALASLTVLPALILAGASRLPLPTTGAGATGTSLSPVSSASA
LGTAAVLAIKMLPIITGMGVAENPTRGGAOVLPGNLPPVVATKSRADLPALAI
AINQVSHRLVEVPGVRKVESAMPAGVFWTDAJSSASAGRIADLQAGAGSFPAVTA
```

```
IKSMKSIIEOMSGAVDOLDSTVNVNTLAGARQOQYLDPMILAAARNLKKNTTELSEYLE
TITHIVGFTNCPDDVLTAMKXVTEPEPDIYVTTGNELSTGADRSATSTQTHALSS
APRWYAKRSLALAOVRSFVPLETTIDAMPQIDQASAMLNKLSADPADTEGGTHLS
RKDLADPSYRHYRESFSSDGTARFLFYSGQIDQLAARQOOLEIAAGKAMKYSGL
VDSQVTGGAQOIAAARADALIHDAVILLVILLVLAIVLALVLAISMRGAVGAIVLAS
YLAALGVSTALMOHLDRELNALVPLSEFVLAISGVYLVAGIAGRIADPATGARS
KGAVSGCAVAPVLAALGVPFCAGLVLSGSGSFVLSQIGTVVVLGSLVITQKRAMLP
TPPKRR"
7769..9801
/gene="mmpL7"
/note="PS00639 Eukaryotic thiol (cysteine) proteases
histidine active site"
Query Match 98.8%; Score 891.8; DB 2; Length 24292;
Best Local Similarity 99.8%; Pred. No. 8.2e-179;
Matches 893; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 gatttcgaacagagactggagcgcggttgcgataacccttcagcggttcgagctgatacgy 60
Db 7110 GATTTCGAACAGAGACTGGAGCGCGGTTGCGATPACCCCTGACGTGCGACGCTATCCG 7169
QY 61 cgaacgctgaatctgycacagagagctgaacgctgtgtgtccacgaggttacccgctggtg 120
Db 7170 CGAACGCTGAATGTGGCACAGAGAGCTGACCGCTTGTTGTTCCACGGGTACCGCGTGTG 7229
QY 121 atctctgctcgcgaaggactcgaagtaactgctgcctcctcctcgcgcgcttgcaagccgag 180
Db 7230 ATCTCTGCTCCCGCAGGACTCGATGACGTGCTGCTTTCTGCGCGCTTCACAGCGCGG 7289
QY 181 cgcatacgccgttcgcttcgcttcgacaaagcgcggttacccagatgaagcttcgattcg 240
Db 7290 CGCATCGCGTCCGCTTTCGCTTCACAAAGCGGCGTTACGATGAACGTTTCCGATTTCG 7349
QY 241 gtaactgaatctgctgcgcgcgcttgccatctcctcctacatactgctgcgcgttggaacgctc 300
Db 7350 GTACTGAGTGAATTCGTCGCGCGGTGGCGCATTCACACTACATCGTCTGCGCTGACGACGTC 7409
QY 301 gtgcaacatgctgycgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 360
Db 7410 GTGCAACATGTTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 7469
QY 361 ctgcagcttggaagctccggaatggtatatacttcaaaagaagaagatatacctataccgcg 420
Db 7470 CTGCAGCTTGGAAGCTCCGATGAGGTATACCTTCAAGAGAGAGATACATCTACCGCG 7529
QY 421 tatttgcaatacactcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 480
Db 7530 TATTTGCAATACACCTCGCGGTCACCCCGCAGCGCGCGCGGTGTGATATGCCATCAG 7589
QY 481 aacgttcgggttaatttcgaacagctgatactgctactcttcgagataccgaagatc 540
Db 7590 AACGTTCCGGTTAAATTTCCAMACAGCTGATGCTGCTACTTTGCGGATPACCGACGGGATP 7649
QY 541 ccacgcgcaaatctcgcaatgctatccgtgctaccccttaccagaagaatggttgta 600
Db 7650 CCACGCGCAAAATTCGCACTGCTATCCTGCTACCCCTTCTACACGACATGGGTTGGTA 7709
QY 601 ataggaaatttcgcacaaatcttggttgatatacccgcggtgctcaccagccgcgctgctc 660
Db 7710 ATAGGAATTTGGGCAACAAATCTTGGGTGTGATACCCCGGCTCTCACACGCGCGGTGTCG 7769
QY 661 ttctctgcaagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 720
Db 7770 TTCTCTGCACGCGCGCGCGCGGTGATGCACTTGATGCGCGACGATTTTACACCGCTTTTTCG 7829
QY 721 gaagcaccgaatttcgcttgaactagcgcgaagaagaacaacgaacagacacatgccc 780
Db 7830 GCAAGCACCAATTTGCTTTGAACCTAGCGGACGACAGACACACGACGACGACATTTGGCC 7889
QY 781 gggcgttgacctcggaacatactgacacatccctcaacgctgtagcgaacggttacaagccgcg 840
Db 7890 GGGCGTGACCTCGGCAACATACATGACATCCCTACGCGGTAGGACGCGGTACAGGCGCG 7949
```

QY 841 acgataacgctctgcgcacgcttctgctcattcgaatgagagaggtga 895
|||||
Db 7950 ACGATCAAGCGCTTGGCGGACCGCTTGTCTGCTCAATCTCGAGAGAGGTGA 8004

RESULT 2
LOCUS MBU75685 1749 bp DNA BCT 19-APR-1997
DEFINITION Mycobacterium bovis acyl-CoA synthase gene, complete cds.
ACCESSION U75685
VERSION U75685.1 GI:1658530
KEYWORDS
SOURCE Mycobacterium bovis.
ORGANISM Mycobacterium bovis
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacterineae; Mycobacteriaceae;
Mycobacterium; Mycobacterium tuberculosis complex.
1 (bases 1 to 1749)
Fitzmaurice, A.M. and Kolattukudy, P.E.
Open reading frame 3, which is adjacent to the mycocerosic acid
synthase gene, is expressed as an acyl coenzyme A synthase in
Mycobacterium bovis BCG
J. Bacteriol. 179 (8), 2608-2615 (1997)
97252490
2 (bases 1 to 1749)
Fitzmaurice, A.
Direct Submission
Submitted (22-OCT-1996) Ann Marie Fitzmaurice, Neurobiotechnology
Center, The Ohio State University, 1060 Carmack Road, Columbus, OH
43210, USA

FEATURES
source
Location/Qualifiers
1..1749
/organism="Mycobacterium bovis"
/strain="BCG"
/db_xref="taxon:1765"
/clone_lib="pyb18 cosmid library"
1..1749
/codon_start=1
/transl_table=11
/product="acyl-CoA synthase"
/protein_id="AAH52538.1"
/db_xref="GI:1658531"

CDS
1..1749
/translation="MAMSYSRLPALRACACLOPHDPAFTMDYEDMDGVAITLTWS
QYRTLVNARELSRCSTGDRVYISAPGLEYVAFALAGRLAVPLSVPGGVT
DEKSDVLSDSVALITLTSADVQVAVRPPGSPSIIEDLIDLDADNGLYFK
DEBYESTAYLOYTSGSTRPAGVYSHQVNRVEDMGYADTGTIDLPNSALVSM
LPEYHMGIVIGICAPILIGYPAVLTSPPVSIORRMHMLASDHFASAPNFAFE
LAARRTDDDMAGRDGNLITLISGSEKQATIKRFARFARNLQERYVIRSYELA
EATVVAISKPGQPEPTVDFTDESISAGAAKPAAGGATSLISYMLRSPVIRIVSD
TCIECDGTVGEIWHGDNVNGYWKRPESERTGKIVTSPGTPGEPWLTGDSG
FYTDEKMFITGRIKDLIVYGRNHSPPDIEETIOETTRGCAAI SVPGDRRTKELVAI
IELKRGSDSDAMARLGAIKREVTSALSSHGLSVADLVVAPGISPIITSGKVARG
ACVEOYRQDQFARLDA"

BASE COUNT 351 a 518 c 517 g 363 t
ORIGIN

Query Match 98.4%; Score 888.6; DB 2; Length 1749;
Best Local Similarity 99.6%; Pred. No. 8.3e-178;
Matches 891; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 181 cgcatacgctgacgcttctggttccaaagcgcggttacgcatgaacgtlccgaltcg 240
|||||
Db 265 CGATCGCCGCGCGGCTTGTGGTTCCAGAGCGGGGCTTACCATGAACTTCGATTGCG 324

QY 241 gtaactgagttatctgtccggttgccatctcaatacgtctcggtgagcagctc 300
|||||
Db 325 GTACTGAGTATTGTCGCGCGGTGCGCATTTCTCAATCAATGTCGCGGAGAGCTC 384

QY 301 gtgcaacatgttcgacgacgacgacgacgacgacgacgacgacgacgacgacgacg 360
|||||
Db 385 GTGCAACATTTGCGCGGCGCGCGGAGATCCCGGCATCATTTATGCAAGTTGATTG 444

QY 361 ctgcatacgttcgacgacgacgacgacgacgacgacgacgacgacgacgacgacg 420
|||||
Db 445 CTCGATCTGACCTCCGAATGGGTATACCTTCAAGAAACAGATATCATCTACCGCG 504

QY 421 tatttgaaatacctccggttcacccgacgacgacgacgacgacgacgacgacgacg 480
|||||
Db 505 TATTTGCAATACCTCCCGGCTCACCCGACGCGCGCTGCGTGTGATGTCCTACAG 564

QY 481 aacgttcggttaatttcgaacagctgattctgctgaacttgcggaatacgaacgagatc 540
|||||
Db 565 AAGCTTGGGTTAATTTCGAACAGCTGATGTCTGGCTACTTTGCGGATACGACGGGATT 624

QY 541 ccacgcgcaaatctcgcaactcgatctcgtgctacaccttcaacagacatgggttgta 600
|||||
Db 625 CCACCGCAAAATTCGCGACCTGATCTGCTGCTACCTTTCACACGACATGAGTTGGTA 684

QY 601 ataggaaattgacgaacaaattctgggttgatgatacccgagtgctacacgacgagtgctg 660
|||||
Db 685 ATAGGAATTTGGCCACCAATTTGGGGTGGATACCCCGGGTGTCAACACCGGATGTCG 744

QY 661 ttctcgaacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacg 720
|||||
Db 745 TTCTCTGAGGCGCGCGCGGTGATGACATTTATGATGACGATTTTTCACGCTTTTGC 804

QY 721 gcaacacgaatttcgaccttgaaactagcggcagaagaacacgacgacgacatggcc 780
|||||
Db 805 GCAGCACCGCAATTTCCCTTTCGTAATAGCGGACGAAGAACAACGACGACGACATGCGCC 864

QY 781 gggcgtgacctggcaacatactgacatctcaagggttagcagaggggtgacagcgccg 840
|||||
Db 865 GGGCGTGAACCTCGGCAACATCTGACCATCTCAGCGGTAGCGAGGGGTACAGCGCGCG 924

QY 841 acgataacgctctgcgcacgcttctgctcattcgaatgagagaggtga 895
|||||
Db 925 ACGATCAAGCGCTTGGCGGACCGCTTGTCTGCTCAATCTCGAGAGAGGTGA 979

RESULT 3
LOCUS MSGMYACSYN/C 9699 bp DNA BCT 26-APR-1993
DEFINITION Mycobacterium bovis mycocerosic acid synthase gene, complete cds.
ACCESSION M95808
VERSION M95808.1 GI:149977
KEYWORDS mycocerosic acid synthase.
SOURCE Mycobacterium tuberculosis (strain BCG, pathovar bovis) DNA.
ORGANISM Mycobacterium tuberculosis
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacterineae; Mycobacteriaceae;
Mycobacterium; Mycobacterium tuberculosis complex.
1 (bases 1 to 9699)
Mathur, M. and Kolattukudy, P.E.
Molecular cloning and sequencing of the gene for mycocerosic acid
synthase, a novel fatty acid elongating multifunctional enzyme,
from Mycobacterium tuberculosis var. bovis Bacillus Calmette-Guerin
J. Biol. Chem. 267, 19388-19395 (1992)
92406887

FEATURES
source
Location/Qualifiers
1..9699
/organism="Mycobacterium tuberculosis"
/strain="BCG"
/db_xref="taxon:1773"

```
CDS
complement(153..1049)
/note="ORF"
/codon_start=1
/transl_table=1
/protein_id="AAA25367.1"
/db_xref="GI:149978"
/translation="MDYEQMDGVAITLMSOLYRRLTNVARELSRSGSGTDRVYIS
PGLEEVYAFGLAAGRILAVPLSPGCGVDERSDSLSSPSVALITSSADVVY
OHVARRGESPSLIEVDLDDLPANGTFEEDEYPSIAYIQYISGSTRTAGVMSH
ONVRFNEQOLMSGYFADTDGIPPNISALVSWLPYHMDGLVIGCAPLIGYPAVLVS
PVSELRQRRMMLMADSFHAFSAAPNFAELARRTTDDMAGRDIGNILITISGSE
RVQAATIKPFADALLASICRRG"
738..1556
/note="ORF"
/codon_start=1
/transl_table=1
/protein_id="AAA25368.1"
/db_xref="GI:149979"
/translation="MLHDVYHGRRCSENGHRRITNOYRIGTFIGNALMNRKHGDA
GLORAEKGDVLESLRTPDHHAIVRGTTAQLPHOICRSPITOLRPGRYRAVYPLV
IHKREGRVMGLQTRRRNAERETLNGITPATLGVYAVAVGCSRAVIAKSHLAS
SAYCRLAMGPSPDTVGCDVAPPRGRRLTOLRHCARAGTENCGATDVLRGRSP
MTAPNWLALPSVEILRMSNSPATSPRLPEQLAPATWGLIPRSEVRLGHEHVCSC"
1741..8073
/codon_start=1
/transl_table=1
/function="fatty acid elongating enzyme"
/evidence="experimental"
/product="mycocerosic acid synthase"
/protein_id="AAA25369.1"
/db_xref="GI:149980"
/translation="MESRVPVAVITGMCCRLPGGINSRDKMESLIRGDDLVTEPRD
RMDDDYDDEPGYVGRSVSRMGCFLLDVACFDDEFREISREKRTSIDPQRLLETTS
WEALIEHGLDPAFLAGSSSTAVFTGLTHEDYVILTTAGLASPVVYGLNNSVSGRI
AHTLGHGPAMTPTACSSGLMAVHLACRSIHDERDLALGGCAVGLLEPAACAASA
QGLMSTGRCHSFADADGFEVRSRGCAMVILKRLPDLADNRNIFAAYRGTATQDGR
TETLTPSEDOVAVYRAALAAAGVOPETVAVYHAGGPTISPIEYRSRATYAGG
TPCALGSAKSMGHSSTASAGVGLIKATLSRHCVPLRLHFNPLPDLSTVEIGLPLV
POAYTPMNGNDHTPKRYAVSSFGMSGTNVATIEAPAEKSADESSDEVEGRFL
MLSTSSDALRQIARQLATVWEHQDCVAASDLVYTLARGRRHPRVRAVVAANPEL
VEIGREVAUSDALDAVAGHGRDPVWVFSQGSQOMAMGTQLASBPFAATTAKLE
PVIAAESGFATVETAAQOVTYIDKQPAVFAVOALAMTEQTYGVRPAVYVSHM
GESAAVYAGALSIEDARVTCRSKLMTRTAGAGAGSVYELPAKOVSELMARGID
VYVSVASPOSTVIGTSDVYRDLIARMEQDWARAVAVVASHSPQVDIIDDLLA
ALADIAETPKVPIYSATLDPREQPVCDGATVYDNELNVTQFAAQAQAEDEGRVF
ADVSPHLTHAVBQOTGRSLDMSVAALAGMRERQPLPGLGLTELHARAGALDYSA
LYPAGRLVDAPLPAMTHARLFIIDGQEOBORGACTITVHLLGSHVRLTEPRHW
OGDVTGYLSWLSHVNVAALPGAAECMAALAAAEFGEAAEVRDITPEOMILLD
EOTPIDAVASIDAGVAVNEVETNRDGETTHATAALARAADDCCPPREGYDITLLOAH
PHAVNGTAMRESFAERGVITGARGLTTLATTPRPRRCRRCRRCSSRAPE
STRCWTLYVSRKASRYSRYGHMPPAYAVGCAQACLRAYRQCPULLHRAVDQGLQRD
PRWEADLDVDEHGTVLLAVGLRMGTSTSRDRDRLYSRLLTLGMOQALPEVGD
GEAGSWLIDTNSVDPFDMASLTJTDALKSHGQGECAALSVDYTPNDQAGLE
KLGSQLRGDDVYVYVGRVGDDESHSLAGREOVRIHLYRTRELAEGBELPLFYV
TROQOIKPHDSEGRANLEOGLRGLRYSSEPMRLTLLIDVENHTDVERVAQOOL
SGSEDEETAMNGDMYVARLTPSPLGHEBERTAYLDDHOGMRQVPRPDQTLTEYV
ASDRYPGPGQOIEYAVSMSSINPADVILARFPIIDDRPQLQMDVYGVVATVAGEV
TGHVGDVGFSEGGCWRTFLTCDANLAVLPGLTDEQATATATAHAAVAGYJNDL
AQIKAGKVLHSAITGGVQAALISIAARKGAEITATGPNPKRMLMAGMEHYDNR
SVEFAEOIRRDITDGYVDVILNSLTGAORAGLELPGFGFAVEIGKADYVGNRLG
PFRFGLTEFYLDLALMSVQDPRVRELAVFLTLDGVTAGFVEIGKADYVGNRLG
RAMSNAHTGKIVLDVPSGRSVAVPEQRIKRRGSIYITGGLGGLGFFRSKLA
AAGCRVITVARSOPNPKAROTIGSLRAAGDIYVEGONIAEPTADPLVSAATAGL
PLRGVLSAAVEDATLNTIDELIDRWSRKYVSGNMLHRAATLGOPLDWCLESSGA
ALIGSPGQAYLAANSWVDFAHMRKQGLVRSIANGAMEVSRATFLADGEMFTM
PEEGAYAFETLVHRDAVSYIPIILGAPMLADLVRSIANGAMEVSRATFLADGEMFTM
ELSLPQDENAGRLRLLEBOASVILRRTIDADSPIFYGDISLGMLEMRHVTETRG
IRLTPKVIATNKTARALLAOYLADTLAEQAAAPAS"
complement(8468..9532)
/note="ORF"
/codon_start=1
/transl_table=1
/protein_id="AAA25370.1"
CDS
```

Db 206 ACAGTACAGCGCTTCGCCGA-CGCTTGCTGCCTCAATCGAGAGAGCGTGA 153

|||||

RESULT 4
MTCY409 30352 bp DNA BCT 17-JUN-1998

LOCUS Mycobacterium tuberculosis H37Rv complete genome: segment 158/162.
DEFINITION 297188 AL123456
ACCESSION 297188.1 GI:3261805
VERSION
KEYWORDS
SOURCE
ORGANISM Mycobacterium tuberculosis.
Bacteria: Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacteriineae; Mycobacteriaceae;
Mycobacterium.

REFERENCE 1 (bases 1 to 30352)
AUTHORS Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C.,
Harris, D., Gordon, S.V., Eigmeier, K., Gas, S., Barry III, C.E.,
Tejeda, F., Baddock, K., Basham, D., Brown, D., Chillingworth, T.,
Comor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S.,
Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J.,
Moule, S., Murphy, L., Oliver, S., Osborne, J., Quail, M.A.,
Rajandream, M.A., Rogers, J., Rutter, S., Seeger, K., Skelton, S.,
Squires, S., Squires, R., Sulston, J.E., Taylor, K., Whitehead, S. and
Barrell, B.G.

TITLE Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence
JOURNAL Nature 393 (6685), 537-544 (1998)
MEDLINE 98295987
REMARK Erratum: [[published erratum appears in Nature 1998 Nov
12:396(6707):190]]
2 (bases 1 to 30352)
TITLE Parkhill, J.
AUTHORS Direct Submission
JOURNAL Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium
tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome
Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique
Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux,
75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk
On Jun 27, 1998 this sequence version replaced gi:2224815.

COMMENT Notes:
Details of M. tuberculosis sequencing at the Sanger Centre are
available on the World Wide Web.
(URL: <http://www.sanger.ac.uk/projects/M.tuberculosis/>) CDS have
been renumbered from the original cosmid submissions but the old
gene designations are in brackets after the new gene numbers.
Gene prediction was based on a Hidden Markov Model of TB genes
implemented in Tbpase (Krogh) supplemented with visual inspection
of positional base preference in codons, especially where there is
an increase in the observed/expected third position G + C.
CAUTION: In some cases we may not have predicted the correct
initiation codon. Where possible we choose an initiation codon
(atg, gtg, or ttg) which is preceded by an upstream ribosome
binding site sequence (optimally 5-13bp before the initiation
codon). If this cannot be identified we choose the most upstream
initiation codon.

FEATURES
source Location/Qualifiers
1..30352
/organism="Mycobacterium tuberculosis"
/strain="H37Rv"
/db_xref="taxon:1773"
misc_feature <1..2270
/note="fragment designated v026. Does not represent a
physical clone"
misc_feature 18..23
/note="Asn1 site, fragments S/G4"
gene 95..1714
/gene="csp"
CDS 95..1714
/gene="csp"
/note="RV3811, (MTV026.16), len: 539, csp, some short
similarity to gp1x66078|CGCOP1G_1 C. glutamicum cop1 gene

forPSI. (657 aa), FASTA scores: opt: 250 z-score: 244.4
E():4.2e-06 39.4% identity in 109 aa overlap. Contains
PS00687Aldehyde dehydrogenases glutamic acid active site"
/codon_start=1
/transl_table=11
/product="csp"
/protein_id="CAB10019.1"
/db_xref="GI:3261806"
/translation="MAATVYIAMIAIRNPASSHSPSPNPNOLANOPILIGCGVTV
RELTDTPPSLVALGDLAGTSARRAKRPDDGKPMVTEYETEPDRPADGDSYEL
GLINPGRSPDPVFGTTTVOVAATRPIDAPITPPAPRPNDLDSGLGRPAKE
QPFQONISAILISPPQAPPTQMTPTAVTMAAGOPAILISRAEMGDSGLRTEYD
RVRRAAVVHTAGSNDYSPLESAGIKAYITVHAKTIGWCDIAYALVADKYGQVEGS
AGGLIKRVGFGHTGGFNRMVGMVGMINDVAPPIOTIRYGRLLGRMLGMDVDR
SMVDDQSASSSTTPPGGAIARLPAFTTRVDVGNDCPGNAAYAVADELRDIAHFND
PPEELIKALEGATITRMQALGMSALCAPTSPEDAADGARIATFAAGANWSPVT
DAOPITGATYEAMASQSYERGPGLPTSAIEQPELQITONFQHGLNFERLGNVTEV
VGITTPLATRPSPGPTVPEHFTLPTHPIT"
392..415
/gene="csp"
/note="PS00687 Aldehyde dehydrogenases glutamic acid
active site"
1859..1862
/note="possible RBS for PGRS, agga"
1868..3382
/gene="pe_pgrs"
1868..3382
/gene="pe_pgrs"
/note="pe_pgrs"
/note="RV3812, (MTV026.17, MTCY409.18c), len: 540. Glycine
rich hypothetical protein belonging to PGRS (pe) family,
similar to many e.g. YM36, MYCTV_010873 hypothetical 53.7
kd protein cy39.36c (558 aa), fasta scores: opt: 327
z-score: 458.5 E(): 2.5e-18, 28.5% identity in 502 aa
overlap, also similar to MTCY7H7B_27; MTCY5_25;
MTCY493_24; MTCY441_4; MTCY39_36; MTCY411_4;
MTCY359_33; MTCY130_10; MTCY98_9 etc."
/codon_start=1
/transl_table=11
/product="pe_pgrs"
/protein_id="CAB10020.1"
/db_xref="GI:3261807"
/translation="MSFVYVPEAAVAAAGDLAAGSTLREATAAAGPTGLAANA
DDVSTAVSOLGFRGVEGQFOTGVNOLAHTTEVRLNRRAGANLMTESANGOLGQIT
EAGQRAVSAAANAARCGATGQIVATATATLSELVYKANSNPPPLRIOTIANQVWQO
IAALANAVQNPALVLANLPAAIDAVDQFLAFNAAYTQITISSIGFADLPATTVG
QGVTSVIAQPKLIGPLGDLFTTMTIPDAQAGFTNLMPSILRDSQNTNVLTLN
GTTISVTAQKPLIGPLGDLFTTMTIPDAQAGFTNLMPSILRDSQNTNVLTLN
PNIOAVASPDIAATTAGTSTFGVPLVPLVIAQAPFALNATSATRIEQQALLGN
YIGAVGALIDAPAHALDGLFNSATVLPDPIILVPTGAPSLPPTVGTGTHLPDGLIIV
PHPVATITSPGCAVPPIRGFTTVYVGFPPGMABDLINLYIPQALAIKRAA"
2170..29622
/organism="Mycobacterium tuberculosis"
/strain="H37Rv"
/db_xref="taxon:1773"
/clone="Y409"
complement(3691..4512)
/gene="RV3813c"
complement(3691..4512)
/gene="RV3813c"
/note="RV3813c, (MTCY409.17), len: 273. Unknown, similar
to many hypothetical proteins eg. YXEH_BACSU P54947
hypothetical 30.2 kd protein in idh-deor (270 aa), fasta
results: opt: 329 z-score: 456.0 E(): 2.2e-18, 32.28
identity in 267 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein RV3813c"
/protein_id="CAB10018.1"
/db_xref="GI:2224832"
/translation="MKPTVPALVACDVDTLLDDGETVTKRTTDAVHAADVAGTHFIL

[illegible][illegible]

/organism="Mycobacterium tuberculosis"
 /db_xref="taxon:1773"
 /clone="Y409"
 BASE COUNT 7322 a 12527 c 13809 g 7663 t
 ORIGIN

Query Match 47.2%; Score 426.2; DB 2; Length 41321;
 Best Local Similarity 67.5%; Pred. No. 1.2e-80;
 Matches 599; Conservative 0; Mismatches 288; Indels 0; Gaps 0;

```

QY 1 gattacgaacagagactcggagacgctgtgacataacctgacgagctgcagactgtatcgg 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 18316 GATTACGAACAGGAGTTCGGAGGCATTAAGTAAGAACGCTGACGTCGTCGACGAGTATCGG 18257

QY 61 cgaagcctgtaattgagacagagctgagagcgtgtgtgttcacagaggtgcgagctg 120
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 18256 CGAACCTTAAGCTTGCAGCAGAAAGTCGCCGCATGCCCAATTGGTGACCGTGACAGTG 18197

QY 121 atctgtctcgaagagactcgaactcgaactcgtctgcctatctcgcgcgtctgcagcgcg 180
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 18196 ATATTGGCCCCACGAAGACCTGATTAATGTTGCTTTCTGGGCGCTTACAGGCCGCT 18137

QY 181 cgcatacgctgcgcgcttcgtgtccacaaagcgagcttacagatgaagcttcgattcgg 240
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 18136 CTATTATGCGGCTTCACCTTTCGCTCCGCTCGGCGCCACGATGACGATGCTTTGACGCG 18077

QY 241 gtaactgaatcgttcgcgcgagctgcgccttactactacgtctgcgcgagcgcgcgc 300
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 18076 GTATGTCGTCACCGCAAAACCAATGTCCTTCTACACATCCCGCATATGGCGCATGTC 18017

QY 301 gtcgaacatgttcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 360
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 18016 GTCCCGCGGCTTACGCGACCGCCGCGATGTCGACGCCCGCAACGCTTGGCGTGATCAA 17957

QY 361 ctgcatactgagacgtcgcgaatggtatcttcaagaagaagagatccatctaccgcg 420
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 17956 CTAGATCTGACCTCGCGATACGATCTAATATTGTGTGACATTTCTCCAAACACCGCA 17897

QY 421 taattgaataacacctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 480
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 17896 TATTTCAGATATCGTCGGGATGACCGCACACCTGCGGTATGATTAATGATTAATCAAG 17837

QY 481 aacgtcgcgttaattcgaacagctgaatgtctgtacttctgcgtatccgataccagcgag 540
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 17836 AATATATTGGCAATTTCCAGACAGATGATTTCCGCTATTTCCGACACCGGACCGCTA 17777

QY 541 ccacgcgcaaatccgcgaactcgtatccctgtactacccttaccagacatgggtttgta 600
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 17776 CCGCCATTGGACCTTTTCAATTATGTCGTGCTACCGTTCTATCATGACATGGGTTGGTT 17717

QY 601 ataggatattgcgacaaatctgggtgataccgcgcgcgcgcgcgcgcgcgcgcgcgcgc 660
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 17716 CTGGGAGTTTGTGCGCGATTAATGATGAGATGCGGCGCTGTCTCACAAGCCGGTGGCG 17657

QY 661 tctctgcagcgccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 720
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 17656 TTTCTCAGACGACCGCGGTGCTGCAATGATGACGACGAGGCGACGCGGCTTTTGG 17597

QY 721 gcaagcgcgaatttcgccttgccttgccttgccttgccttgccttgccttgccttgc 780
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 17596 GCGGCGACCGAATCTCGCTTGCATGACGCGACGCAAAAGCAATGATGACGACTTGGCC 17537

QY 781 ggcgcgtgacatcgcgaacatactgacatccatccagcgttagcagaggttacagcgcg 840
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 17536 GGGCTCGACCTTGAGAGGATCAAAACATCTCTGCGGCGAGTAAAGGGTGACATCCGGCG 17477

QY 841 acgatcaagcgttcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 887
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 17476 ACCCTCAACGCGCTTGTGACGCGGTTAGCCGTTTCAATCTTCGAGA 17430
  
```

RESULT 6

MTV045
 LOCUS 11042 bp DNA BCT 17-JUN-1998
 DEFINITION Mycobacterium tuberculosis H37Rv complete genome; segment 67/162.
 ACCESSION AL022000 AL123456
 VERSION AL022000.1 GI:3261541

KEYWORDS
 SOURCE Mycobacterium tuberculosis.
 ORGANISM Mycobacterium tuberculosis.
 Bacteria: Firmicutes: Actinobacteria: Actinobacteridae:
 Actinomycetales: Corynebacteriineae: Mycobacteriaceae;
 Mycobacterium.

REFERENCE
 AUTHORS

1 (bases 1 to 11042)
 Cole,S.T., Brosch,R., Parkhill,J., Garnier,T., Churcher,C.,
 Harris,D., Gordon,S.V., Eigemeier,K., Gas,S., Barry III,C.E.,
 Tekala,F., Badcock,K., Basham,D., Brown,D., Chillingworth,T.,
 Connor,R., Davies,R., Devlin,K., Felkell,T., Gentles,S.,
 Hamlin,N., Holroyd,S., Hornsby,T., Jagels,K., Krogh,A., McLeau,J.,
 Moule,S., Murphy,L., Oliver,S., Osborne,J., Quail,M.A.,
 Rajandream,M.A., Rogers,J., Rutter,S., Seeger,K., Skellern,S.,
 Squares,S., Squires,R., Sulston,J.E., Taylor,K., Whitehead,S. and
 Barrell,B.G.

Deciphering the biology of Mycobacterium tuberculosis from the
 complete genome sequence
 Nature 393 (6685), 537-544 (1998)
 98295987
 Erratum:[[published erratum appears in Nature 1998 Nov
 12:396(6707):190]]
 2 (bases 1 to 11042)

REFERENCE
 AUTHORS
 TITLE
 JOURNAL

Direct Submission
 Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium
 tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome
 Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique
 Molculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux,
 75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk
 On Jun 27, 1998 this sequence version replaced gi:2918949.

Notes:
 Details of M. tuberculosis sequencing at the Sanger Centre are
 available on the World Wide Web.
 (URL, <http://www.sanger.ac.uk/Projects/M.tuberculosis/>) CDS have
 been renumbered from the original cosmid submissions but the old
 gene designations are in brackets after the new gene numbers.

Gene prediction in rlpA (Krogh) supplemented with visual inspection
 of positional base preference in codons, especially where there is
 an increase in the observed/expected third position G + C.
 CAUTION: In some cases we may not have predicted the correct
 initiation codon. Where possible we choose an initiation codon
 (atg, gtg, or ttg) which is preceded by an upstream ribosome
 binding site sequence (optimally 5-13bp before the initiation
 codon). If this cannot be identified we choose the most upstream
 initiation codon.

FEATURES

source

source

gene

CDS

1..11042
 Location/Qualifiers
 /organism="Mycobacterium tuberculosis"
 /strain="H37Rv"
 /db_xref="taxon:1773"
 <1..1595
 /organism="Mycobacterium tuberculosis"
 /strain="H37Rv"
 /db_xref="taxon:1773"
 /clone="Y1965"
 complement(23..6349)
 /gene="pks5"
 complement(23..6349)
 /gene="pks5"
 /note="RV1527c, (MTV045.01c-MTCY1965.01), len: 2108, pks5,
 probable polyketide synthase, highly similar to many, e.g.
 MCAS_MycBO_Q02251 mycocerosic acid synthase from
 Mycobacterium bovis (2110 aa), fasta scores: opt: 6270
 z-score: 7088.8 E(): 0, 63.6% identity in 2126 aa overlap"
 /codon_start=1
 /transl_table=11

D	7700	GCTATCGGGTTTCGCGTCCGGTTCCACATGCCGGCCCAACAGATAGAGATTTCG	7759
OY	241	gtaactgaatgatcgcgcggtaggcacattccatacatcgctctgcgtlgagcaagtc	300
D	7760	GTGCTAACCGACACTTCCGCCCGTGATCTGCAGCAGTCCGGGGCGTTGACGATGTC	7819
OY	301	gtgcaacaatgttgcgcggcggcgaggaaaccgcccaaatlatsgaagtgtttg	360
D	7820	ACAGATCCGCTCACACCACAGCCAGGCCAGTCCGACCATTAATCGTTGAGCTTGATTG	7879
OY	361	cfcgaactcgcgaacctcccaa-----tggtatacctccaagaagaagcag	405
D	7880	CTGGACTTAGATTTGTCGGCAGCGCTCCCGAAGCCCTGGCGCGGCCCAACCGCAGAGAT	7939
OY	406	tatccatctaccgcgatalttgcaatatcacctcogggtccaacccgcgaocccgcttgagtg	465
D	7940	ACGCCCGGAACCCGGTATTTCATAATACCTTCGGGATCCACCCTGACGCCGCGCGGTGTC	7999
OY	466	gtgatgtcccatacgaacagcttcgggttaatttgcgaagactgatatgtctactttgcg	525
D	8000	ATGGCTCTGAACAACAAATAGTCTTGCCCAATTTTCGAGCAGATCGTGGCCGCACTTCTTTGGG	8059
OY	526	gataccgcgaagatltccacccgcacaatlccgcgaactgtatccgcgtactacattcacac	585
D	8060	CCCAGAGGGGGCGTGTCCCGCCGAGCTCACTGTGCTGTGGCTGCCGCTGTATCCAC	8119
OY	586	gacatgggtttgtgaatagaatttgcgcacaatlcttgggtgataccccgcggtgtctc	645
D	8120	GACATGCGCTTCTTATTATAGCGCGCATATCCGATCTCGCGGGGTGTACCCACCGTTTG	8179
OY	646	accagcccggtgtcgttgttcgcagcgcgcggcgcggtgagatgtaactgtahgscagcgt	705
D	8180	ACGATCCGGGGGTTCTTACACGGCCGGCTGTGATGATPAACACTGCTGCACATAAC	8239
OY	706	tttccagccttttcgtgcagcacgcgaatlctgcctttaaactlagcgcgacgaagaacacc	765
D	8240	GGTGCACGATTTGGCGCAGCAGCAATTTGCTTGCATTTGGCGGTGGCTAAGACGTGA	8299
OY	766	gacgaagacatgaccgcgggcgttgacctgcgcgaacatactgacalctcctcagcggtagcag	825
D	8300	GACGACGACATGAGCAGGACTTGACCTCGCGCGGTGCACACCACTCCMAACGCGACGAG	8359
OY	826	cgggtgacagcgcgcgaatcagaagcgttgcgcgcgaacgccttgccttcaatct	881
D	8360	CGAGTACACCCGCGACCCCTCAAACGATTTGCTGAACGATTGGCGCGCTTAAATTAT	8415
SOURCE	NCBI		
ORGANISM	Mycobacterium tuberculosis.		
DEFINITION	Mycobacterium tuberculosis H37Rv complete genome; segment 66/162.		
VERSION	277826.1 GI:3261623		
KEYWORDS			
REFERENCE	1 (bases 1 to 11430)		
AUTHORS	Cole,S.T., Brosch,R., Parkhill,J., Garnier,T., Churcher,C., Harris,D., Gordon,S.V., Eigmeier,K., Gas,S., Barry III,C.E., Tekala,F., Badcock,K., Basham,D., Brown,D., Chillingworth,T., Connor,R., Davies,R., Devlin,K., Feltwell,T., Gentles,S., Hamlin,N., Holroyd,S., Hornsby,T., Jorgels,K., Krogh,A., McLean,J., Moule,S., Murphy,L., Oliver,S., Osborne,J., Quail,M.A., Rajandream,M.A., Rogers,J., Rutter,S., Seeger,K., Skelton,S., Squares,S., Squires,R., Sulston,J.E., Taylor,K., Whitehead,S. and Barrrell,B.G.		
TITLE	Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence		
JOURNAL	Nature 393 (6685), 537-544 (1998)		

```

MEDLINE      98295987
REMARK       Erratum:[published erratum appears in Nature 1998 Nov
              12:396(6707):190]]
REFERENCE    2 (bases 1 to 11450)
AUTHORS      Parkhill,J.
TITLE        Direct Submission
JOURNAL      Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium
              tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome
              Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique
              Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux,
              75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk
              On Jun 27, 1998 this sequence version replaced gj:1477429.
COMMENT      Notes:
              Details of M. tuberculosis sequencing at the Sanger Centre are
              available on the World Wide Web.
              (URL, http://www.sanger.ac.uk/Projects/M.tuberculosis/) CDS have
              been renumbered from the original cosmid submissions but the old
              gene designations are in brackets after the new gene numbers.
              Gene prediction was based on a Hidden Markov Model of TB genes
              implemented in TParse (Krogh) supplemented with visual inspection
              of positional base preference in codons, especially where there is
              an increase in the observed/expected third position G + C.
              CAUTION: In some cases we may not have predicted the correct
              initiation codon. Where possible we choose an initiation codon
              (atg, gtg, or ttg) which is preceded by an upstream ribosome
              binding site sequence (optimally 5-13bp before the initiation
              codon). If this cannot be identified we choose the most upstream
              initiation codon.
FEATURES
source       Location/Qualifiers
              1..11430
               /organism="Mycobacterium tuberculosis"
               /strain="H37Rv"
               /db_xref="taxon:1773"
               /clone="Y1965"
               /cdate="Y1965"
               /gene="Rv1519"
               /size=81..350
               /note="Rv1519, (MTCY1965.09c), len: 89. Unknown, high
               similarity to C-terminus of MTCY78.26, 58.1% identity in
               74aa overlap."
               /trans_start=1
               /codon_start=1
               /product="hypothetical protein Rv1519"
               /protein_id="CAB01397.1"
               /db_xref="GI:1477438"
               /db_xref="SWISS-PROT:Q50588"
               /translation="MRGCGACGCVLCANSGRRPRALCTAVATRTLSLTNML
               VESADLTVEEDICSRIVSLPVHMHAIADARVYAFGEGLANG"
               /size=376..1416
               /gene="Rv1520"
               /size=376..1416
               /note="Rv1520, (MTCY1965.08c), len: 346. Probable
               transferase, similar to e.g. TR034224 (EMBL:U07786) SUGAR
               TRANSFERASE from Vibrio cholerae (337 aa), fasta scores:
               opt: 214 z-score: 215.5 E(): 8.4e-05, 25.9% identity in,
               212 aa overlap"
               /trans_start=1
               /codon_start=1
               /trans_table=11
               /product="hypothetical protein Rv1520"
               /protein_id="CAB01396.1"
               /db_xref="GI:1477437"
               /db_xref="SWISS-PROT:Q50587"
               /translation="MSIVSISYNOEETIEALDGFAPORTPEFVPIITADDSSTQATP
               RTIGGEAARYPOLFRILKOTNIGVIANRKDVLSARKEYLALCEDDYWTPIRLKSK
               QNVLDRHPETTYVCFPRVRYIEDGKDSFEPPLSRRLSDALLARNLIGVNSVY
               RRPQSDDDIPANVPIDWYLVHRAVAGEIAMLPEYMAVYRRHAGHIMSAITDRKRF
               WERTGSMATLEAMLDLVGHGREPAIVGEVSAMVREIGKPGCGGALLKTSIAD
               HPMRTLSLOHRAQCPWRRFRKRRLSTELSLAALAVATRRRLLEGROGYSRTTSP
               TGRGRNVRCSHA"
               /size=1650..3401
               /gene="fadD25"

```


Db	2274	AAGTCGCCCGCCAGGCTCCACCGCTGGTGTGTGTGGTCCGTTCTATCCAGCATGGGT	2333
OY	595	ttgtgaatagaatttgcaccaaattcttggatggatccccgcggttgcacagccgcg	654
Db	2334	TTTCGCTTGGGAGCTATATTGCCATTTCGGCTGGGATCCCGCGGTGTGACAGCCCG	2393
OY	655	gtgtcttctctcagcgcgccgcggtggatgtaacttgaatgagcagatttcaagcc	714
Db	2394	ATCGGTTTCTTCACAGCGCCCGCTGGCTGGATACAGATGTTGGCAAGCAACTCTTGGC	2453
OY	715	ttttgcagacacccgaatttcgctttgaactacagagggcaggaagaacacagcagac	774
Db	2454	TTTTACCGCGCCGCCAAGCTCCGATTCGATTCGGGTCTGTGTAAGACCAAAAGCAGAGAC	2513
OY	775	atggcgcggcgctgacactcgcacaactacatcctcctcagcgtgagtagcagcggtacag	834
Db	2514	ATGAGAGGCGCTCGATCTCGGTGGCGCTACAGGCGATCTCTAAGGCGCAGCAAGGGTGCAG	2573
OY	835	gcccgcagcatcaagcgtcttcgcagccgctttgctgcctcaatctcagagagaggtg	894
Db	2574	CCGGTAGCGCTGAACGGCTTCATCGACGCGGTTCGCGCCGCTCAATCTTGACCCCAAGCGC	2633
OY	895	aaa aag 897	
Db	2634	ATA 2636	
RESULT	8		
LOCUS	MSGY456	37316 bp	DNA
DEFINITION	Mycobacterium tuberculosis sequence from clone y456.		
ACCESSION	AD000001		
VERSION	AD000001.1	GI:1702974	
KEYWORDS			
SOURCE			
ORGANISM	Mycobacterium tuberculosis (clone: y456) ds-DNA.		
REFERENCE	Mycobacterium tuberculosis		
AUTHORS	Bacteria; Firmicutes; Actinobacteria; Actinobacteriidae;		
JOURNAL	Actinomycetales; Corynebacterineae; Mycobacteriaceae;		
COMMENT	Mycobacterium tuberculosis complex.		
FEATURES	1 (bases 1 to 37316)		
Source	Location/Qualifiers		
	1..37316		
	/organism="Mycobacterium tuberculosis"		
	/db_xref="taxon:1773"		
	/clone="y456"		
BASE COUNT	7226 a 12014 c 11317 g 6759 t		
ORIGIN			
Query Match	42.6%; Score 385; DB 2; Length 37316;		
Best Local Similarity	65.0%; Pred. No. 6.2e-72;		
Matches	587; Conservative 0; Mismatches 310; Indels 6; Gaps 1		
OY	1	gattcgaacaggaactggagcggcgttcgcataaacctcgaacgcggttcgacgtctatcg	60
Db	987	GATTTCAGAGCGGCTCTCGGATGTTGAAGAAACTCATCGTGGCGCAGTTATATCGG	928
OY	61	cgaacgcgaatgtggcacaagagcgtgaacccgttgggttccacaggttgaacgcgtggtg	120
Db	927	CGAAGCGTTAACTTCGCCGACACAGCTTAAGAAATGAGGCTGACCGGCATGGGATTA	868
OY	121	atctctgtccgcaggaactcgaactacgtcgtccctactcggcgcgttcgacgagccgg	180
Db	867	ATTCTGGCGCCACAAGCCTCGACATATGTCGTTAGCTTATTGGCCCTCGCTGACGCGGA	808
OY	181	cgcacgcgcgttcgcttccggttccacaagcgcgcttaccgatatgaacttccgattc	240

[illegible]

TITLE	Direct Submission
JOURNAL	Submitted (01-MAR-1994) Department of Genetics, Harvard Medical School, 200 Longwood Avenue, Boston MA 02115
COMMENT	On Mar 31, 1994 this sequence version replaced gi:414216. This sequence data was produced by the Genome Sequencing Center located at Collaborative Research Incorporated (1365 Main St., Waltham MA, 02159). 617-487-7979). Please contact Doug Smith (smith@cric.com). The annotation should be considered preliminary and incomplete.
FEATURES	Location/Qualifiers
source	1. .41171
CDS	/organism="Mycobacterium leprae" /db_xref="taxon:1769" complement(771..1655) /note="probable enoyl-CoA hydratase (similar to Arthrobacter sp. fcb8); Lepb1170_C2_224" /codon_start=1 /transl_table=11 /product="fcb8" /protein_id="AAI17070.1" /db_xref="gi:466794" /translation="MSQDASCTIAELPYSVTDLVLDPPREVALITLNPRGMS MALDLMKSLKOVLRITVDHSRVVVLGAGRGCSGADOKETAIVPQVEGLTOPVRA LRAMELEEVILALRLHQPIAINGPAIGGGLALADYRVASTRAYRAAGINN GLSASELGLSYLPRAGVSGSAFELMLSGRDVGAEEAOIGIVSYRVSDRLDTCYS IAARWATFSRGSTELTKRALMGIDLAASLDKHMOSESLAQFIALHTSNFEBAARCT EKRPTVVDARGCATSPG" 2640..2861 /note="possible regulatory protein or antibiotic resistance protein; Lepb1170_F3_104" /codon_start=1 /transl_table=11 /product="rpl" /protein_id="AAI17079.1" /db_xref="gi:466803" /translation="MLTLSEPDVLFDESTNDVNTLTATEDEFLISMAVTLSRITG YLHSATEKQLTSDIRPSARLPKPDNRG" complement(3581..3982) /note="similar to stillbene synthetase; Lepb1170_C3_257" /codon_start=1 /transl_table=11 /product="sts" /protein_id="AAI17074.1" /db_xref="gi:466798" /translation="MRQAMEHQDVTITTTITGAPSLDARCAEQGLRIDVARM PLFGACVTGQGLGRKHEDIVWCSGRGCGIWSARSPITPSNQLCRAMCGDGSVRA PHRRRTDLCCQCAHPRFAQPSLPRNRCISMA" complement(7087..7341) /note="probable ketoacyl synthase (similar to polyketide synthases); Lepb1170_C1_189" /codon_start=1 /transl_table=11 /product="kas" /protein_id="AAI17064.1" /db_xref="gi:466788" /translation="MSYFPGMGTRVTVDSSOSSIVAVYLTGCSLRTGASPLAVLRM HPSQGERNRYIEKVCWGIIECWPHLSGVALGWTCTTTE" complement(8884..9066) /note="as-associated gene (similar to orf-III located 5' of BCG mas gene; Lepb1170_C1_188" /codon_start=1 /transl_table=11 /product="mas8" /protein_id="AAI17063.1" /db_xref="gi:466787" /translation="MPPELVNVOGSLQLRPSQCFADTRPLLVGGEAFIWLQA GTFQERKDGRIHGL" 8893..10011 /note="as-associated gene (similar to orf-III located 5' of BCG mas gene; Lepb1170_F1_11" /codon_start=1 /transl_table=11 /product="masC"
CDS	/protein_id="AAI17075.1" /db_xref="gi:466799" /translation="WQNSSVLSFLRERAGLPDDEAFSTFYEQNMAGVTKTLWAO L YORTLVNVAHERLRHGSIRDAVAVILAPQGLDTIAFLGMQAGFIAPVLSVQASDHE RVGAVLADTSPSVLITTSVADALAKYVSDSDTVPVAILVDSDSNLDENKSNIRLT DAPSTAYLOYTSGRTRLEPTGVYMRINLMVNDQOLMAYFAPTNVAPLADLTIVSLP FYHWDGLVAVAPLIGGWRSELSPISEFLORPARMDIOMATSSHPFAGNPAFELA ARSDIADIAGLDIAAGCGITISGSRRIHPALINRPSDEAFINFRDMMILPSYGLAEC TVYASALREARORSFISNREAVGHRHETAHGPARHC" 10038..10628 /note="probable CoA ligase; similar to potato 4-coumarate--CoA ligase; Lepb1170_F3_111" /codon_start=1 /transl_table=11 /product="pc1a" /protein_id="AAI17080.1" /db_xref="gi:466804" /translation="MRIYSDPTCIECPTRGVYELWLHGDNVAEGYMRKPEPTORFGG KLANSPGTPGPMLETRGDLGISEDELFTVGRMKDLIVYGRNHYPEDISTVOEIT GGRVAATSVPPDETPEKLTITTEVKKRGSDAEMAKVAVANNNTTAISKSHGLNAD LVLPDGSITPTTSGKVRRTACVEQYRRQFSRLDG" 10638..11350 /codon_start=1 /transl_table=11 /product="Lepb1170_F2_64" /protein_id="AAI17078.1" /db_xref="gi:466802" /translation="WMTITLMAIAISLEPPRIGMTVMINRPPITQLVFLCSGFT MGMTGVFVLEVFERRRLMASMQLLPKVOILIGVILALVAVALVYOVCISSPEAESP VDSAGPSPKPMKAPRLARLNGSLDNLVAVAGVIGLIPVDVILAAVILVYAGAA TTYOGALIMFNVVAFALIEIPLAAYLAPDTRAMVTALNNMISRRLEAVATLLACV SCILAVGIAICL" 11415..12590 /codon_start=1 /transl_table=11 /product="Lepb1170_F3_112" /protein_id="AAI17081.1" /db_xref="gi:466805" /translation="WASRLGAIARAKSLPGLCLAPPGYSEKNKNEKKRLATTLA LTVGVSGFSTVASSEGEPEWLPSPVPVPEVENSAPKIVIALGGRPPTFMDVYTR AGDEFFPDVNRKLLIDYPARAPFRVPPFLVPGPRDEVTIGBAIIVATKNLQALHRTG EPAAVVGLSQGLADIDYEOELADTPAPPDQLETFNFGDPSGVHGFGKSVLASIFR PGDYPLIDYTMPQMSODXSDNRVAVYDGLSEPPDADNLALQNLCFAGCAISHP SGFENPEDEPPQNI RTTYNSRGAKTTTILIVNHLPLTLRIYLGMSAIVDOIDAVL QPKIDAAIAYNDPNLKPISYDPVNGMDPIAGIDAEIADSLTNVFAQLRSILPPEPG" complement(13018..16044) /note="rpl; Lepb1170_C1_181" /codon_start=1 /transl_table=11 /product="transport protein (similarity to antibiotic transport protein actII-3 from S.coelicolor)" /protein_id="AAI17062.1" /db_xref="gi:466786" /translation="MHPGCAITGELIFESRLGDIYTRMPVWVIGCWTLLALMPMTVP S LTELTORHPVILVYPDAPSSVAARKISQAEHDESENVILYLLPDGKLGADDEVYK TIVNRLRNDTDVYVLODPLSTPRLHELTVKDKOKALPLVYVLGELGTSKSYVAC VAGIVKQTLBESTAGSSSLKANLRGPASTVADLPDAGARISITELVIALVLTIIIMIT YRNPITMLPLITITGASIMTQAQVAVSGVSLAGVIAASATVGTITFELGMSFTFIRKVS TV ISRHVDYIRMGSGADGACAVROALISLQVIVASATVGTITFELGMSFTFIRKVS TV GPALAIGIAVAFIAAVTLPALVYLAGRGVAPPRDAGAFMRGTGVAIRVAYL SASVILIVLALCASIVRENYDRKQIIPASPESSYVALSHSPVGOATREYVILLOS PHDRTPRALADMAELAQROVQICIAIVRYVTRPGRLRETSAITYOAGWVGQDGS ASHLIGESTGDLNRLASGAGLADLGVRLQVGOVAVGISGLDNLNFAQRMGFGDSK TLGELDTAGKLVSSMRALGNMFGINEFTTMDIMWGAVVIALISSVLCQDNPICAD A RAOPHKLITJASEDGTDLNIAHLMKQIGSTQSGQITGAVSGLEKTLFVAVNTSLRSLG DNPVYMRSKMIGLQNGVNDLASAGRIADGVAIVDORTGTMTGLARSAPFLAMEGOD ASQPSMAGENVPOLNTEDEPKLVQATISPDGSHVRYTQDOLNPSSAMDOVNTI LNVATGADPNNTLSDASTYLSGYTLPEDPDDYDRDROLIVTMIVVALLITMALLR STVAPTIYGVSVIYSALSLCVAVFOVLLRQOLHNSVPCLAIVYLVACADYNML ASRLDESSSHGLRASVIRTVRSTGIVTAAULIFRASFFGLLSIATVVOAGFVLGS GILDTFIVRTIVPAVALRLRASWMPARWGNNA" complement(16051..17466) /note="weak match to surfactin synthase; Lepb1170_C1_180"
CDS	

```

/codon_start=1
/transl_table=11
/product="sus"
/protein_id="AA117061.1"
/db_xref="GI:466785"
/translation="MOVGLTGLTLDLMADMRACKTITSMOPTNCEKVSQPVSV
AVMOACHIRYVQKAKGLDYSLMTYSCQQLQCCDRAINITYNMLRHNDYRSE
EYIDEGEILRHLCIDPADIEFPIEHQULSLAOTRELAVTDPDLGWCFQVQIAE
EHFTFYASIDHVDAMIVGLTMEFMMMSALVSGAPLEPAGSYDDFCRROHRE
TSMILTAEPOIRSMTOFAPIENEGSPFDPPLDPLEPDTLRIIDVLTLMDEQOTDRE
AACTVAGARVGVGLACGLAEYELTGADYVPLGPDRTLRIDITDVLTOGMPLGDI
TPYIAGSSFGDARAANONCEDTVHLAEVYDGVVLAELAPVHKPRPPEPYNIDGT
APLSVLTAGLDLNLNGVSDGRXYTOMSTIYVREOETAIVAMFDPNPEAGSVART
LETKSVFECVAGSHMRNVA"
complement(1/547..23903)
/note="mycocerosic acid synthase; lepbl170_C2_209"
/codon_start=1
/transl_table=11
/product="masa"
/protein_id="AA117069.1"
/db_xref="GI:466793"
/translation="MRAANTPVAIVIGMACRLPGIDSPORLMEALLRQDVLTEIK
ERMADLYDDEPVGPRSVSRMGAFLLDVGDPDPEPTEPEALAIIDPOHLLLET
SMADIHAGIDPATTLGASATGVVGLTHQDYELSDCGATEBPFTGTSSNFSAGR
VATTLHGPAVTVDRACSSGLMAVHQARSLDCCGSDALAGVAVVLEPRKSVGS
LQGLMSPTEKCAFDVAADGFVSEGVVLLKRMPEARDGRIDLAVLRGTAAANDG
HTVHIAPSEBOAIYVHRAVAGVDPVTLGLEAHGTGPVGDPIEYASLAAYGT
ACRCALASTKTNQNGHSTGSLPKMKAILALOHGVVPMKHLFTRLDPLAETTELF
VPOVNTPMVVDQOPRRAVSSYSGMSTNVHLLBQAPNSGSDSPAVSKVNPILFP
LSATSVOLRATARMADWDEHSGSTKRTGLMDLGYTISRRCGRPVPRMVLASSFA
ELSRGLRETAGSETPYOSAVGODDRGPVWFSGSSOMGMELELRAKEVEFAALIA
MEPVILEESGFSTVETALVAEKTVIDRQVPTIFAIOVLAALMKSYGVPVGVHIS
LGEAAVAAGALSCEDJRVLCRSLQSGRAMASVELPQSGVSELSIRLS
DVLAVVASPTSTVVGDTQAIHRHLAAMQAOVMVREVAVDVASHTPQVEPLIDLLV

```

Query Match 41.2%; Score 371.8; DB 73; Length 41171;
 Best Local Similarity 63.7%; Pred. No. 3,76-69;
 Matches 565; Conservative 0; Mismatches 322; Indels 0; Gaps 0;

QY 1 gatacgaagaagactggaagcagctgtgacataacactgacgcagcgtgacgtgacg 60
 Db 8971 GACTACGACGAGGATTTGGCGGCTGTCGCGCAAAACACTGACTTGGCGCACCTGTATCA 9030
 QY 61 cgaacgctgaagtgtgacacagagcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 120
 Db 9031 CGAACCCTTGGAACCTTGGCGACGAGCTCAGCGGACCGCTTCGATAGAGACCGGCGAGTG 9090
 QY 121 atctctgctcgaagagacacgtgacgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 180
 Db 9091 ATATTGGCTCCCAAGGGGCTGGACTATATCATGCGCTTCTGGAGCTATGCAAGGCTGGC 9150
 QY 181 cgaatcgccgtgacgcttctgtgttccacaagcggtgtacgaatgaagcttcgattcg 240
 Db 9151 TTATATGCGCGTTCGGTGTAGTACCTCAGGCGCGCTCCATATGACGAACTGTGCGAGCT 9210
 QY 241 gactcgaagtatctgtcgcgcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 300
 Db 9211 GTCTTGGCCGACACATACCTCGTGTCTGTACACAGCTGTGCGGTGTCGATGCGATC 9270
 QY 301 gtgcacaatgttcgagcgccgagcggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 360
 Db 9271 GCGAAATATCTTACCACTCTGTACACCGACCGTCCCGCGATCTCGAAGTGTGACTCA 9330
 QY 361 ctgcacatgcagctccgaatgt 420
 Db 9331 CCAATCTTACGCTCGAGATTAATCAATATAGACTACCGATGCGCGGACGACACGCG 9390
 QY 421 tatttgaatatacctctcgggtgtccaccgcaagccgcgtgtgtgtgtgtgtgtgtgtgtgt 480
 Db 9391 TACTTACAGTACACCTGCGGCTTCGACCGGCTTCGAACCGGGTCAAGTGTGACGACGCGG 9450
 QY 481 aaagtcgggttaatttcgaacagcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 540

```

Db 9451 AACCTTATGGTGAATTCACAGTGTATGGCTGACTTCCGTCACCTAACGCTGTG 9510
QY 541 ccacgcgcaaatctcgactcgtatccctgtaccccttccacgaatggtgtgta 600
Db 9511 GCGGACATTCATCTCACCATCTGCTGCTGGCTTCTACACGATATGCGTGTGTT 9570
QY 601 atagaatttcgacccaattctggtgtgatacccgcggtgtctacacagccggtg 660
Db 9571 CTCGGCTCGTCGACCAAACTTCTGGCGGTGGCGGACGACGACTTACGATCCGATCTCG 9630
QY 661 ttctctgaagccgcccggcggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 720
Db 9631 TTTTTCGACGCTCCAGCTGATCCAGCTGATCCAGCTGATCCAGCTGATCCAGCTGATCC 9690
QY 721 gcaacgacgaatttcgacttcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 780
Db 9691 GCAAGACCGAATCTTGGCTTCTGATTTGGCTTGTGACGCGGACATCAGCAGATATAGCC 9750
QY 781 gggcgtagcttcggaacatactacacatccctcagcgtgtgtgtgtgtgtgtgtgtgtgt 840
Db 9751 GGCCTCGATCTCGGCGCTTGCAGGCGATCATGACGGGAGTGAAGCAATTCACCCGCC 9810
QY 841 acgataagcgttcgcccgcgcttgcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 887
Db 9811 ACGCTAATCGGTTCTGATCGTTTGCAGATCAATTCACCCGCA 9857

```

RESULT 10
 MTUV05/c 37840 bp DNA BCT 17-JUN-1998
 LOCUS Mycobacterium tuberculosis H37Rv complete genome; segment 51/162.
 DEFINITION AL010186 AL123456
 ACCESSION AL010186.1 GI:3261493
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Mycobacterium tuberculosis.
 Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 Actinomycetales; Corynebacterineae; Mycobacteriaceae;
 Mycobacterium; Mycobacterium tuberculosis complex.
 1 (bases 1 to 37840)
 Cole,S.T., Brosch,R., Parkhill,J., Garnier,T., Churcher,C.,
 Harris,D., Gordon,S.V., Eigmeier,K., Gas,S., Barry III,C.E.,
 Tekala,F., Badcock,K., Basham,D., Brown,D., Chillingworth,T.,
 Connor,R., Davies,R., Devlin,K., Fellwell,T., Gentles,S.,
 Hamlin,N., Holroyd,S., Hornsby,T., Jagels,K., Krogh,A., McLean,J.,
 Moule,S., Murphy,L., Oliver,S., Osborne,J., Quail,M.A.,
 Rajandream,M.A., Rogers,J., Rutter,S., Seeger,K., Skelton,S.,
 Squares,S., Squires,R., Sulston,J.E., Taylor,K., Whitehead,S. and
 Barrell,B.G.
 Deciphering the biology of Mycobacterium tuberculosis from the
 complete genome sequence
 Nature 393 (6685), 537-544 (1998)
 Erratum: [[published erratum appears in Nature 1998 Nov
 12;396(6707):190]]
 2 (bases 1 to 37840)
 Parkhill,J.
 Direct Submission
 Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium
 tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome
 Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique
 Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux,
 75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk
 On Jun 27, 1998 this sequence version replaced gi:2695948.
 Notes:
 Details of M. tuberculosis sequencing at the Sanger Centre are
 available on the World Wide Web.
 (URL, http://www.sanger.ac.uk/Projects/M_tuberculosis/) CDS have
 been renumbered from the original cosmid submissions but the old
 gene designations are in brackets after the new gene numbers.
 gene prediction was based on a Hidden Markov Model of TB genes
 implemented in TParse (Krogh) supplemented with visual inspection
 of positional base preference in codons, especially where there is

an increase in the observed/expected third position G + C.
CAUTION: In some cases we may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtc, or ttc) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.

FEATURES
Location/Qualifiers

source

1..37840
/organism="Mycobacterium tuberculosis"
/strain="H3/Rv"
/db_xref="taxon:1773"

misc_feature

<1..36758
/note="fragment designated v005. Does not represent a physical clone"
70..1977

gene

CDS

/gene="lpqW"
/note="Rv1166, (MTV005.02), len: 670. Possible membrane protein, almost identical in part to M. tuberculosis gene fragment TR:G238465 (EMBL:AF009358) ORFA2-898 (FRAGMENT) (59aa)(93.9% identity in 49 aa overlap) (Nano F.E., Doran J.L., Treit J.D., Moran A.J., Identification of Mycobacterium tuberculosis peptides that stimulate immune human peripheral blood monocytes. Unpublished). Contains possible N-terminal signal sequence and appropriately positioned PS00013 Prokaryotic membrane lipoprotein lipid attachment site. Also contains"
/codon_start=1
/transl_table=11
/product="lpqW"
/protein_id="CA15843.1"
/db_xref="GI:2695950"

misc_feature

/db_xref="SPTREMBL:O50422"
/translation="MGVSPVRRVCVTGALVALACVLAGCTVSPPPAPSTDPTRS
TTPPRPTQIIMGIDWIGPGENPHLIDLSFVNAISALVPSAFRPIIDPPTGS
RWMNDPLIVSADVYNNHPTVYTKIRPEAMQDNAPIAADFWYIMQMTQPVVD
PAGHYLTIVSDEBGRQAVVTFAPQPYAMRELFTDILPAHIVKDIPEGFASGLRAL
PVYGGQFVENIDPQRDEILARNDRWGSPSPKPIILERRGAPALADSVNRSDTQ
VAAVHGSAFAAQLAISPDVTRATVTPRVMPFTIRANVPKLAIDQVKALIGLDVD
LLAAVAGACTDVTVLDAQAIKSPDPGVVTPAPMSSAAALGLEASCFQVDTSTV
SPAPSPSTTSVSTGPEVYRGRISKDEGLTVIGVAAANDPSTVAANAOLR
DVGLAATVLAIDPVTLYHDALNDNRVDAIVGRQGNLTLASRYCPRALQATTV
AANAFTTAPSPAPIGTPPSAAPTATPPPTAPRPSDPCALVAPSNLTGICDSIQSN
IDAAIATGKINIDVITVAVEPRLMNMTVLPIIDTTIVAAGPSVQVNSLSCAVPVGI
GDAGQWVKTGQ"
121..153
/gene="lpqW"
/note="PS00013 Prokaryotic membrane lipoprotein lipid attachment site"
complement(2005..2610)
/gene="Rv1167c"
complement(2005..2610)
/gene="Rv1167c"
/note="Rv1167c, (MTV005.03c), len: 301. Probable transcriptional regulator, similar to several eg.
TR:D102772 (EMBL:D85417) hemr from Propionibacterium freudenreichii (243 aa), fasta scores: opt: 268 z-score: 416.1 E(): 5.4e-16, 35.9% identity in 198 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv1167c"
/protein_id="CA15844.1"
/db_xref="GI:2695951"

CDS

/db_xref="SPTREMBL:O50423"
/translation="MTVASPAKANYRRRGVLEALYDATALAELESAGYGLTMEGI
AARQGTGAALRRPAKRELYLAOVYALPVPKPADRSARENLVAFVFNCELTA
GKTALPSMEIVSOLHEPELRAIFINSWAPLRLVESILQAGVRSGEIDPATLTPTMT
ARIGPALIHQHLFTGSPDPDRQLTRIDAMILTTGERRES"
complement(2682..3722)
/gene="ppe"
complement(2682..3722)

CDS

/gene="ppe"
/note="Rv1168c, (MTV005.04c), len: 346. Member of the M. tuberculosis PPE family, similar to many eg. TR:E332769 (EMBL:D98268) MTC1125.27C (385 aa), fasta scores: opt: 504 z-score: 780.2 E(): 0, 36.6% identity in 388 aa overlap"
/codon_start=1
/transl_table=11
/product="ppe"
/protein_id="CA15845.1"
/db_xref="SPTREMBL:O50424"
/db_xref="SPTREMBL:O50424"
/translation="MDFTIPPEPENSILNIGSARPELYAANMKNLISNELSASREF
SEINCLITSMRCPSTMAAVALPRAATVITPASTAEVADHISVNAKAYEANAHAH
PLVYETNRRLRLATLTNITGIRTPALFDALDAQVWSODGEMANLITATMAAAR
LTPESPAPALNPGALALYELIGSSETSVSFAAPATKNSPKLMTLLTGTYPLTA
ARISIPVEVYLAPEGSMGCMGLMRSITPTLKGPLELLPNAVPAVASATLGA
DTIGLSVPPSVNADKSIPLAKAVPTAPGPGSGTSAOGLASLAGAVAVART
RSGVILRSPNAG"
complement(3740..4042)
/gene="ppe"
complement(3740..4042)
/gene="ppe"
/note="Rv1169c, (MTV005.05c), len: 100. similar to the N-terminal region of the M. tuberculosis PE family of proteins eg. TR:O05297 (EMBL:D93777) MTC1364.07 (99 aa), fasta scores: opt: 209 z-score: 407.6 E(): 1.6e-15, 37.4% identity in 99 aa overlap. Also similar to the N-terminus of M. tuberculosis TR:P77909 (EMBL:U76006) ESTERASE/LIPASE (EC 3.1.1.3) (TRIACYLGLYCEROL LIPASE) (437 aa), fasta scores: opt: 193z-score: 381.8 E(): 4.4e-14, 37.2% identity in 94 aa overlap. Contains a helix-turn-helix motif from aa 88-109(+2.76 SD)
/codon_start=1
/transl_table=11
/product="ppe"
/protein_id="CA15846.1"
/db_xref="GI:2695953"
/db_xref="SPTREMBL:O50425"
/translation="WSEVTRPDSTIGETAAINHEIGVTMSAHDGVTPLITNVESPAH
DLVSLTSMLEFSMGELYKAIARQAHVHESFVQLQTSKTSYWLTELANNAGYST"
complement(4051..4056)
/note="possible RBS upstream of Rv1169c"
4222..5133
/gene="Rv1170"
4222..5133
/gene="Rv1170"
/note="Rv1170, (MTV005.06), len: 303. similar to Streptomyces lincolnensis lmbe gene TR:O54358 (EMBL:X79146) from lincomycin production gene cluster (270 aa), fasta scores: opt: 308 z-score: 409.5 E(): 1.2e-15, 32.0% identity in 278aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv1170"
/protein_id="CA15847.1"
/db_xref="GI:2695954"
/db_xref="SPTREMBL:O50426"
/translation="MSEPTRLFIYAHNDDSELSNGATIAHYTSKGAOVHVCTLGE
EGEVIGDWAQDLTADHADQIGSYRIGETLAALRALGAYPIYLGAGRWRSQMGATD
QRSORRFADADPROGVATVATIRELRPHAVYTDNPGSGYCPDHVHTHTYTAAVAA
AGVSGGTADHGDPTWPKFYTWYVGLSALISGARALYPDDLRPEWVTPRDETFVGA
SDDGIDAVEADEQARAKVAALAHAVQVVVGPGRAAALSNLALPTLADENHYLA
GGSAGARDERGMEYDLAGLGTASGT"
5303..5665
/gene="Rv1171"
5303..5665
/gene="Rv1171"
/note="Rv1171, (MTV005.07), len: 120. Unknown hydrophobic protein"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv1171"
/protein_id="CA15848.1"


```

/db_xref="GI:2695955"
/db_xref="SPTREMBL:O50427"
/translation="MYVLLALIVDGVSVSALAGALMPRTIGSAPFPLISGLVNA
LVMAAARMTSSRVALLPLMAMLLTVAAANSFGSPDVLVIGGGLLVYGLVPAVNA
VPPAVLMRRNRVADSG"
complement(5673..6599)
/gene="pe"
/cds complement(5673..6599)
/note="RV1172c, (MTV005.08c), len: 308. N-terminus is
Query Match 40.7%; Score 367.4; DB 2; Length 37840;
Best Local Similarity 63.9%; Pred. No. 3.2e-68;
Matches 573; Conservative 0; Mismatches 321; Indels 3; Gaps 1;
QY 1 gattacgaacagagcgtgagcagcgttgcgataacccgacgcgtgcacgtgtatcg 60
31352 GACTACGACAGAGATTTGGCGCGCATCACCAAAACGCTCACGTGCTCAGAGGTTTCGA 31293
61 cgaacgctgaatgtgacagagagctgagccgtgtgttccacggtgtgacgcgtgtg 120
31292 CGGACCGCATCGTTGCTCAGAGATCAGGCGCATTCACAGCAGCGAGAGCCGTCG 31233
QY 121 atctctgctccgacgagcgtgagtaagtcgtgctcattcgcgcggttgcagccgag 180
31232 ATTCTGGCTCCGAGGCGCTGCTACATCCGCGCTTCTAGGTTGATGTCAGCGCGGA 31173
QY 181 cgaacgctgagcgttgcgttgcgttccacagagcggttaccgataagcgttcgattcg 240
31172 GCGATACGAGGTTCCGTTGCTCAGAGATCAGGCGCATTCACAGCAGAGCGTACGCGC 31113
QY 241 gtactgagtgatctgctgcgcgtgtgcatcttcacatcgtctgcgtgtgacgacgtc 300
31112 GTGCTTCCGACGCGTGTACCTTCGTCATCCTGACACATCTGCGTGCAGAGAGCGT 31053
Db 301 gtgcaacatgttgcgcgcgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 360
31052 GCCGAGACATTCACCGCGCGCAACACAGCAACTAGGCGCATTCGAAATGATTCGA 30993
QY 361 ctgcactgagcgttgcgttgcgttccacagagcggtatccatccatccatccatccatcc 420
30992 TTGGACTTTCAGCG---GGAAATTCACCGAGCTTTGGGTTGAAGGATTCGCCAAGCGCGCG 30936
QY 421 tattgcaatacacctccgcgcgttccacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 480
30935 TACTTGCAGATACCTCTGCTGCTCAGCGCGCTCGGCTGGGATCATCTCCATCGG 30876
481 aagcttgggttaatttcgaacagctgagtgtctgctacttgcgtatccgataccgagggatc 540
30875 AACCTTAGGGAATTCACACAGCTAGTGTGCAACTACTTCGGCGACCGCAACGGAGTGTG 30816
QY 541 ccaacgcgaatctccgcgcgttccacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 600
30815 GCGGCACTGACACTACATCAGTGTGCTGCGCTGCGCTTTCACACGACATGGCGTGTGTC 30756
QY 601 ataggaaattgcgcacaatctggtgtgataccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 660
30755 CTGGGATCATCGACCGCATCTGGCGGCTACCGGAGTGAACCTACAGCCCGCTGGCG 30696
QY 661 ttctctgagagccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 720
30695 TTCTGCAAGGCGGCGCGCTGCTACATTCGCTGCGCAACGGAGTCTCTTCTGTGTCG 30636
QY 721 gcagcagcaatctgcgccttgaactgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 780
30635 GCGGCAACGGAATTTCCGCTTCGATTTGGCGTCCGCAAGACAGATGCGGACATAGAA 30576
QY 781 gggcgtagcctgcgcacaatactgacatccatccatccatccatccatccatccatccatccatcc 840
30575 GGCCTCACTTGGGCAACGCTGCGGCAATCAGCAGTGTGCTGAACGATGTCACCCCAAC 30516
Db 841 acgataaagcgttgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 897

```

```

Db 30515 ACGCTAGTAGATTCGCAACCGGTTTGACCGTACATTTCCGCGAAGCATGATA 30459
RESULT 11
LOCUS MTCY349/c
DEFINITION Mycobacterium tuberculosis H37Rv complete genome; segment 131/162.
ACCESSION 283018 AL123456
VERSION 283018.1 GI:3261671
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
1 (bases 1 to 43523)
Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C.,
Harris, D., Gordon, S.V., Eigmler, K., Gas, S., Barry III, C.E.,
Tekala, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T.,
Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S.,
Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J.,
Moule, S., Murphy, L., Oliver, S., Osborne, J., Quail, M.A.,
Rajandream, M.A., Rogers, J., Rutter, S., Seeger, K., Skelton, S.,
Squires, S., Squires, R., Sulston, J.E., Taylor, K., Whitehead, S. and
Barrell, B.G.
Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence
Nature 393 (6685), 537-544 (1998)
Erratum: [[published erratum appears in Nature 1998 Nov
12:396(6707):190]]
2 (bases 1 to 43523)
Parkhill, J.
Direct Submission
Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium
tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome
Trust Genome Campus, Hinxton, Cambridge CB10 1SA unite de Genetique
Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux,
75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk
On Jun 27, 1998 this sequence version replaced gi:1694844.
Notes:
Details of M. tuberculosis sequencing at the Sanger Centre are
available on the World Wide Web.
(URL: http://www.sanger.ac.uk/Projects/M.tuberculosis/) CDS have
been renumbered from the original cosmid submissions but the old
gene designations are in brackets after the new gene numbers.
Gene prediction was based on a Hidden Markov Model of TB genes
implemented in Tlpase (Krogh) supplemented with visual inspection
of positional base preference in codons, especially where there is
an increase in the observed/expected third position G + C.
CAUTION: In some cases we may not have predicted the correct
initiation codon. Where possible we choose an initiation codon
(atg, gtg, or tgg) which is preceded by an upstream ribosome
binding site sequence (optimally 5-13bp before the initiation
codon). If this cannot be identified we choose the most upstream
initiation codon.
location/Qualifiers
1. 43523
/organism="Mycobacterium tuberculosis"
/strain="H37Rv"
/db_xref="taxon:1773"
/clone="Y349"
/complement(26..1885)
/gene="fadD29"
/complement(26..1885)
/gene="fadD29"
/note="RV2950c, (MTCY349.40), len: 619 aa. fadD29,
Probable Acyl-CoA Synthetase, similar to various
mycobacterial enzymes believed to be involved in
polyketide or fatty acid synthesis such as G466799 MASC.
FASTA best: Y061_MYCTU Q10976 hypothetical 67.9 kd protein
cy338.19, (626 aa), fasta scores, opt: 2417, E(): 0,
(50.4% identity in 618 aa overlap), Identical to G560511
FEATURES
source
gene
CDS

```


DB 1576 AATTGGCCCCCGAGGGCCTTGAAATACATCATCGCTTTATGGGTGTGGCAGCGTGGT 1517
QY 181 GCGATGCGCGTGGCGCTTGGTTCGACAAAGGCGGTTACGATACGATCGTTCG 240
DB 1516 GTCAATGGCGTCTCTTCCGGGCGACAAATTCGATTCACGACGAATATCTTCG 1457
QY 241 GACTAGAGTATCGTCCGCTGCGATCTCAGTACGATCTGCTCGTGGAGAGT 300
DB 1456 GCATTACGAGACTCTGCGCTCATTTATTTACGACCGTCTGATTTATGACGAGT 1397
QY 301 -----gtcaacatgttcgcgcgccgagggaatcccgcatcaatcga 351
DB 1396 ACACATATGCGCGACATCGTCCGCGCACAGGGCGAGTCCGCCATTCGTCGTGCT 1337
QY 352 GTGTATGTCGATCTGAGCGTCCGATGGTATACCTTAAAGAAGAGATGCA 411
DB 1336 GTGAGCGCTCTGGACTTGGATTCACAGCGCTCGATCCGATCCGATTCGAGCGTCCG 1277
QY 412 TCTACCGTATTTGTAATACACCTCCGAGTCCACCGCGACCGCGCTGCGTGTGATG 471
DB 1276 AGCAGCGATATCTACAGTACACATCCGGTTCGACCGCGCGCGCGCTGTTCCTC 1217
QY 472 TCCCATCAAGAGTCCGAGTAAATTCGACAGCTGATGCTGCTGCTGCTGCTGCTGCTG 531
DB 1216 TCGACAGAAATGCTATCACCAACTGCTGATACAGCTAATGCTCGATTCGATTCG 1157
QY 532 GACGAGTATCCACCGCAATCCGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 591
DB 1156 GAAAGGTTCCG-----TCACTCTCCGTAATCTGCTGCTGCTGCTGCTGCTGCTG 1106
QY 592 GATTGTAATGAGTATTCGCAACCAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 651
DB 1105 GGGTGTGATCTGGGCTATCTCCCGATGATCAATCAAGACCGCGGTGTGATGATG 1046
QY 652 CCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 711
DB 1045 CCGATGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 986
QY 712 GCTTTGCGAGCGAGTATTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 771
DB 985 CAGATTTCCAGCGCGCAATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 926
QY 772 GACATGCGCGCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 831
DB 925 GATATGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 866
QY 832 CAGCGCGAGTATTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 891
DB 865 AACGTCGCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 806
QY 892 g 892
DB 805 G 805

RESULT 12
MTU00024 31175 bp BCT 29-SEP-1994
LOCUS MTU00024 31175 bp BCT 29-SEP-1994
DEFINITION Mycobacterium tuberculosis cosmid tbc2.
ACCESSION U00024
VERSION U00024.1 GI:560506
KEYWORDS
SOURCE Mycobacterium tuberculosis.
ORGANISM Mycobacterium tuberculosis
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacterineae; Mycobacteriaceae;
Mycobacterium; Mycobacterium tuberculosis complex.
1 (bases 1 to 31175)
AUTHORS Smith,D.R.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 31175)
AUTHORS Robison,K.

TITLE Direct Submission
JOURNAL Submitted (29-SEP-1994) Department of Genetics, Harvard Medical School, 200 Longwood Avenue, Boston MA 02115
COMMENT On Oct 27, 1994 this sequence version replaced gi:414230 gi:414230. This sequence data was produced by the Genome Sequencing Center located at Genome Therapeutics Inc. (formerly Collaborative Research Inc.) (100 Beaver St., Waltham MA, 02154 617-893-5007). Please contact Doug Smith (smith@er.cic.com) for further information. The annotation should be considered preliminary and incomplete.

FEATURES
source
CDS
CDS
CDS
CDS

Location/Qualifiers
1..31175
/organism="Mycobacterium tuberculosis"
/db_xref="taxon:1773"
<3..902
/codon_start=1
/transl_table=1
/product="u0002a"
/protein_id="AA050950.1"
/db_xref="GI:560529"
/translation="SPRCARLAFGSRRLACRVRVAGVALTAANTSRQIEHVRISPT
VADSKRRPRPDLKSDSGFRGTYOIGCTTYVAVVAVLVETIVTSRDKKGVAGC
DAVRVTSKLVTPGTSMKAVSTEDLFCAGTIFERGETVSKLADICAVADY
TMVALDLSANQYSSRAAAAYCVADSEIEFRFHALLFSKDIQAPLEKDFPDNA
RLIEFLARAGVGVKVPDCINSKYIEKVDGLAAAVNHATPTVRVNGTEYEMSTPAL
VAKIEIYADVDPGIDSAATATS"
908..1540
/codon_start=1
/transl_table=1
/product="u0002b"
/protein_id="AA050949.1"
/db_xref="GI:560528"
/translation="VWAARPAERSGDPAVAVRVPSAMVLLIGVIGLFPASMTLVEK
VRLIDPIYVPCSNVNPVYSGSVSTPQASLGPNNPLICPCMKVAVAVIATLLVVASIV
VPLRWYNTGLAVGLIVGAVPYHMLITFOSVIRIGALCPCKMYAVAVIATLLVVASIV
EPRMKNRSGSQRVGARLLYQWRMSLATLMTFTVLLIMVRWDIWSITLI"
1565..4939
/codon_start=1
/transl_table=1
/product="pyc"
/protein_id="AA050948.1"
/db_xref="GI:560527"
/translation="MFSKVLVANKGEIARAEAAVETGCTVAVPYEDRNSQRHLK
ADESVIDIGHPVHAIVSYDEIVATARRAGADATPGGISEMNDIAACAAAGIS
FVGPSEVLELAGNKSRAIAAREAGLPLMSASASVDELISVAGMPPLFVAV
AGGGRGMRVRVDIALPEALIEARERESAFGDPVYLEOVINPHHEVOITLADNL
GDVHILYERDCSVQRHOKVIELAPHLDALRKMYCDANAFAFHIGYSCATVEF
LIDERGEVTEFMNPRVOEHTVTEITDVDLVASQLRIAAGETLEQGLROEDTAPH
GAALOCRTTTEDPANGFRPTAGSARCDPAPVAVSAMAAPTMRNQPVLRLHAQAD
LSGRDLPVAVSARRAIAEFIRGVSTNIPFLOAVLDLDPDFAGVTTVSFIDERPOL
TARASADRGTKLINFADVTYNNPYSRSRSTIYPDDKLPLDLRAAPAGSFORVYL
GPEGRARLRRESAAVGVDTTPRDAHQSLARVTSGLSRPAVPLARTMPLDLSVFC
WGGATYDVALRFLKEDPWERLATTLRAAMNITCLOMLGRNVTGTPPEIYTSFVQ
EATATGIDIFRIFDALNNIESMRAIDAVRETGSAIAEAMCYTGDITDPCQQLTLD
YLLKLAEOIVDAGAHVLAIKDMAGLRPPAQRVLSALRSRFDLPVHLTHDTPGGL
ASYVAAHMGADAVDGAAPLAGTTSOPALSSIYVAAHTEYDGLSISAVCALPFW
BALRKVVAIPRESGLPGTGRVYHHEIPGOLSINLQOALAIIGIDRFEIEKAVGAD
RYLGLRVYTPTSKYVGDALALVAGVSADEFAADPRAPGIPESVGLFLRLEDGP
GMPPEPLRTAALAGRGARPTQLAADDEIALSSVGAQRQATLNLRLPSPKNEH
REAYGDTQSLSNPFYGLRQEEHRYLLEVEVLLIGLEALSEPDDEGMRTVMKILN
COLRVLVDRSIIASVPAEAKEDRGNGHIIAPAGVTVGVCGEVRVAGQGTATAT
EAMKMEAPITAVAGTVERVAVASDIAEMRAETCMW"
5057..5704
/codon_start=1
/transl_table=1
/product="u0002d"
/protein_id="AA050947.1"
/db_xref="GI:560526"
/translation="MQMGYRLAGDEAMRGYRPLARSGALDPDRRCRRRRRTAVP
PRGTPPTDRVRESLENTVTAARDLTGLAVDLTAGSGLGLEALSRAVDSADVAD
QSAVAVIARNIEALIGATLRGAVVAAGVATLSPVLDVADPVPVWSDVADVAIL
AALGTNGWTRGCTVAVVERATTCAPLTWPEGRMRPQVYVGTDRLEALRLEFAN"
5790..6275

```
/codon_start=1
/transl_table=11
/product="u0002e"
/protein_id="AA50946.1"
/db_xref="GI:560525"
/translation="MTGAVCPGSPDPVTLGVHDIIFERRAAQFDEYVVAALVNPACTGM
FDLDERIMAKVESTHLPLNLRVOYHGLVDFVRSCTGATLVKGLRTGDEYELQMA
QMNKHIAVGDTFFVATAPRYSFVSSSLAKEVAMLGDSSELLPEPVNRRLRDLNTER
T"
complement(7144..8076)
/codon_start=1
/transl_table=11
/product="pur"
/protein_id="AA50945.1"
/db_xref="GI:560524"
/translation="MGKGSMTAHATPNEPDYPPPGGPPPPADIGRLRLNCRDPRGII
AAVSTFLARAGANIISLDHSTAGEGTFLORAIPLHGLTAAVDELORDFGSVADK
FGIDYFLRAEAKPKRVAIMASTEDHCLDLMRNRGELKSVYVVAIANHPDLAAHV
PEGVFTHIPATRDTRTEAEQROQLSGNDIVLARYMQLISGLEAIGCLINI
HHSFLPAFTGAAPYORAREGKLTIGATAHVTPVLDGPIIIBODYVRVDTHTVDDL
VRGADYERANLSNAVLMHCCDRVIVHNNOTIV"
complement(8149..9369)
/codon_start=1
/transl_table=11
/product="u0002f"
/protein_id="AA50944.1"
/db_xref="GI:560523"
/translation="MTSTKVEDRVTAAVLAIGHALATASMTWEILMALILGALSA
VVOAVNRRSTIVLLGDPRPTLVIAVGLAASSCSYAVALARSLFRKANFTAM
AFEIGNLTIVVELGIILALMGQFTAAFEVGGPIMILVLAFLRLEFGARLIDARE
OAEGLAGSMEGHAMDMSTIKREGSPRLSPGPTSIAHVEWEMALILDLIIGL
LIAIAIAWDESPFOSPLANHPMANSVMPPIGTIVATSPVCSIGNVLAFLVN
GGISFGCVIAETPADLILPLINIRYITGKRMVLVLGTFYASVAVGVIIEILFGT
TNLIPSRQSAVTMAETISWNTYTNLVIPLVIAALVVRFTSGGLPMLRMGGSPDA
PHDHRHDHLGH"
9306..10433
/codon_start=1
/transl_table=11
/product="u0002ja"
/protein_id="AA50943.1"
/db_xref="GI:560522"
/translation="MARSLPALPLSPPLMLRSALHHIIPGGIVTGVGNDRADV
PLANVSPHSCRGPRVRSVCYATASRMGCPVASEVRGOAISTTPDAAPGLAARR
RLFPAVAVTAAHVPRPALAOSLDPSRYEHPKCDPRYONOLPLPRPHNAITPIS
BRFGNLTGCRFTAMRTLRKIVEADLVLEADIVAGDRLISLSARAGIGPYIA
IANAVSWPYAQRRPLPDVIMTRLEGVAKLYRLERPLFLALQCPMLMNRHGL
SSLGWNLCRITFDGDNHLYADVPELMEPTYDLPAHNEGLGIPMLSPAKRPPTWDSLPT
DRPIVYATLGTSGGRNLIQVINALAEIPLVIAATAAAVT"
10439..10846
/codon_start=1
/transl_table=11
/product="u0002h"
/protein_id="AA50942.1"
/db_xref="GI:560521"
/translation="MPANAFVADYLPGEAAARSAVVVNGSGSLTTOOALVAGPVIG
VAGNLDHLMNEAVERACAGVLLPTEBLKSGRVAGAWQVYSREYFTSRATRCRLR
TRPRMIPACGECSPAHAGPEAPDMVGOLVTST"
11766..12503
/codon_start=1
/transl_table=11
/product="u0002jb"
/protein_id="AA50941.1"
/db_xref="GI:560520"
/translation="MGLWMSRFSVLGQILGVRLVASFAQLFYRPSDAVAEEYHKM
YGNLWVTKTYMGINCMKSVSDMMNOETISELOPSLVEFGTGYGSAVYFANIKR
QIQGPFKVLTVNHNKALDPRAREPDVLEFESSSTDPALAEQIQRKLNKPYGKIFAI
LSDHSMNHVLAEMKILRPLLSAGDYLVEDSNINGNHVPVLPFGPGPYEALIEAEDFE
PNYKHAERENKRGWTSAPNGFLIRN"
12604..13995
/codon_start=1
/transl_table=11
/product="u0002ka"
/protein_id="AA50940.1"
```

```
/db_xref="GI:560519"
/translation="MEETSVAGDPPDAGTSTAPNAPEPVARKKILFNGEANTLAH
VVRPVLARSLDPSRTEVHFACDPRENNLILPLPSPHRIHVSEEVLLKIAQRLP
YNTRLRKRYIAADRKLINELIAPDVVGNRSLISVSNLAQIPIYIAIANWASQARR
REFLPDVWTRFEFGRVPSILRYLRPLIFALYCLPLMWLRKRGISLSDGLCRIFT
DGDYTLVADPELVEPVTYLPANHHYILGVLMSPPVKPPTMMHSI.PTORPIYATLGSS
GKNILQVLAADLDLPVTYTAARAGRHLKNVANAENVADYLYLGEAAALAVYLCH
GGSFTQOALAGVPIVIGLPSNMDQHLMEALERAGAGVLLRTERLANEGVAAAKOV
LSGAEFRQARRLPKPSDQTLPDSRSTSKVRCSSAGCCPRYRLANCCPVEHILIRER
VAVONAEVPRLS"
complement(14306..15133)
/codon_start=1
/transl_table=11
/product="u0002kb"
/protein_id="AA50939.1"
/db_xref="GI:560518"
/translation="MVQKRYAGTAAANTKKVAMAPMFSIIPTLVAAVLPACIDS
IARQCGFEVLVWGSSTDETLDIANTFAENIGERLIHRDTCGVYDANRQVDA
TGYMLPLFGADSDILEADTLTARVAAFGEHEPSLVYGDVIMRSTNFRMGAFDLRL
LFRNICHQALFYRRGLGTIGPYNLRIYVLADWDFNRCFSNPLATRYMHVVASV
NEFGLSNTIVDKFEFLRKLPMSTRIGRLVIVLRMPKVISRAMVMTVISMRRR"
complement(15203..15934)
/codon_start=1
/transl_table=11
/product="u0002kb"
/protein_id="AA50938.1"
/db_xref="GI:560517"
/translation="MKSILKARFTARSAAFEVSRYSERDLKHQFVKOLKSRVDVYF
DVANGSCGYAAGLRAAVKGRIVSEFELSGFTLIESKASTDPAMDOROLHGDGFT
VTINAGAGQSSSVLPMWKSHONAPRANVYGOEASIHRLDSVAREPLFGMNVATL
KVYQGPBKQYLAGKSTIIDHCYGMQLESLFPLREGMLIPALDLYVSLGFTLTG
LLPCFIDANNRMLQADGIFREDD"
17208..17933
/codon_start=1
/transl_table=11
/product="u0002j"
/protein_id="AA50937.1"
/db_xref="GI:560516"
/translation="MRLPGMLRPTAERHFHSIFYLRHNRQEHIALTLGLDLGNKSVL
```

Query Match 35.1%; Score 317.2; DB 2; Length 31175;

Best Local Similarity 61.7%; Pred. No. 1.3e-57;

Matches 549; Conservative 0; Mismatches 323; Indels 18; Gaps 2;

```
Qy 1 gattacgaacagagactggagcggcttgatgaaccctgaacgagctgcgaactgacg 60
|| || || || || || || || || || || || || || || || || || || || ||
Db 22325 GACATGACACAGATCCGCGTGGCTTACACCGAACCAGTGGCTGGCAATCCACGG 22184

Qy 61 cgaacgcggaatgctgacacaggaagctgagccgctgtggttccaacgggtgaacggcgtagtg 120
|| || || || || || || || || || || || || || || || || || || || ||
Db 22385 CGAGCGATGATCGTGGCAGAGAGACTCTGGATATACGGCTCAAGCGGTGATCGGCTCGCG 22444

Qy 121 atctgtgctcgaaggaactgagtaagctgctgcatactcggcgcttcgaacggcg 180
|| || || || || || || || || || || || || || || || || || || || ||
Db 22445 ATATTGGCCCCCGAGGCGCTTGATATCATTCAGCTTTATGAGGCTGTCTGACGAGCTGT 22504

Qy 181 cgcacgcgcgtgacgcttcggttccacaagcgagcggttacgcagatgaacgcttcgacatcg 240
|| || || || || || || || || || || || || || || || || || || || ||
Db 22505 CTCATTGGCGGTCTCTCTCCGGTGGCACCAATTCGCGCATTCACGACGAGCAAGATATCTTCG 22564

Qy 241 gtacgaatgattcgtgcgcggtgagccatttcacatacatgcttcgacggtgagcagctc 300
|| || || || || || || || || || || || || || || || || || || || ||
Db 22565 GCATTACGAGACTCTGCGCCCTCCATTTATCTTACGACCTCGTGGTATTATGACGAGCTC 22624

Qy 301 -----gtccaacagttgagcgcgcgcgcgaggaatcccgccaatcaatcga 351
|| || || || || || || || || || || || || || || || || || || || ||
Db 22625 ACGACATATGGCGCCACATGCTGCGCCCGCACAGGCGCAGTCCGCCCATGCTGCTGCTGCT 22684

Qy 352 gtgattgtctcgatcgtgagcgtccgaatggtgataccttcaagaagaaagcagatlatcca 411
|| || || || || || || || || || || || || || || || || || || || ||
Db 22685 GTGACGCTCTTGAGACTGATCCAGAGCGCTGCATTCGAGACTGGTTCGAGAGCTCGC 22744

Qy 412 tctacgcgtatttgcataataccttcggttccaacccgcaacgccccgctgagctgtagttagt 471
```

DB 22745 AGCAGACATATCTACAGTACACATCCGGTTCACCCGCGCGCGCGGTGTTCCTC 22804
472 tccacacagaagctcgggttaatttcgaacagctgattgctgacttggagatcc 531
DB 22805 TCCACACAGATATCTACACCACTGCTGACACTAATGTCATTCATTCAGGCGGATTCG 22864
532 gacggattccacacgcaaatccgactcgtatcctggtaccccttaccagacatg 591
DB 22865 GAAAGAGTTCG-----TCACCTCCCGTATCGTGGCGCTTCATCAGACATG 22915
592 ggtttgtaataagaaattcgacacaaattctggtgagatacccggtgtccacagc 651
DB 22916 GGGTTATGCTGGGATCATTTCTCCGATGATCAATCAACACACCGGCTGTGATGAGT 22975
652 ccggtgtcgtcctcgacgcccgcgcggtgatatgaactgagtcgaacagatttacc 711
DB 22976 CCGATGGCATTTCTGCAAGCGCGCGCGCTGATGCAATTAATGCAACACCGCGCA 23035
712 gaccttcgacgacgcaaatctcgaccttgaactagcggcagcaagaacacgacgac 771
2306 CAGATTTCAGCGCCACGATTTCCGTTTCGATGAGTGGCGGTACGAGAACATCCGAGAC 23095
772 gacatgcccggcgctgactcggcaacatactgacacacacacacacacacacacac 831
DB 23096 GATATGCTGAGACTTACCTCGCGGATGTCGAGCATGCTCAACCGCGCGAGAGTA 23155
832 caggcgacgacgac 881
DB 23156 AACGTGCGACGCTGCGGCGCTTCACGAGAGGTTGCGCCCTTCAACCT 23205

RESULT 13

MSGY2 31176 bp DNA BCT 03-DEC-1996
LOCUS Mycobacterium tuberculosis sequence from clone y2.
DEFINITION AD000009
ACCESSION AD000009.1 GI:1702967
VERSION
KEYWORDS
SOURCE Mycobacterium tuberculosis (clone: y2) ds-DNA.
ORGANISM Mycobacterium tuberculosis
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacteriales; Mycobacteriales;
Mycobacterium; Mycobacterium tuberculosis complex.
REFERENCE 1 (bases 1 to 31176)
AUTHORS Du, L.
TITLE Direct Submission
JOURNAL Submitted (11-OCT-1996) L.Du, Genome Therapeutics Corporation, 100
Beaver Street, Waltham, MA, USA, 02154 du@critc.com
COMMENT GSDB:S:1004709
FEATURES
Location/Qualifiers
1..31176
/organism="Mycobacterium tuberculosis"
/db_xref="taxon:1773"
/clone="y2"
BASE COUNT 5417 a 9198 c 10570 g 5991 t
ORIGIN

Query Match 35.1%; Score 317.2; DB 2; Length 31176;
Best Local Similarity 61.7%; Pred. No. 1.3e-57;
Matches 549; Conservative 0; Mismatches 323; Indels 18; Gaps 2;

QY 1 gattcgacacgactcggagcgctgtgataacacactgacgagcgatgatactg 60
DB 22325 GACTATGACACGAGCGGCTGCTTCACCGAAGCGTGAAGTCCACCGG 22384
QY 61 cgaacgctgaagctgacagagctgacgctgtgtgttccagcggtgacacgctg 120
DB 22385 CGAGGATATGCTGCGACAGGAGGAGCTCTGATATACGCTCAAGCGTATCGGCGG 22444
QY 121 atctgtctcgcagagactcagatcgtcgtcgtcctatcgcgagcggttgcagcgcg 180

DB 22445 ATATGCCCCCGAGGCGCTTGATATCATCATTCGCTTATGGGTGCTGACGGCTGT 22504
QY 181 cgcatacgcgttcgcgcttcttggttccaaagcgcgcttaccagatgaacgttcgatac 240
DB 22505 CTGATTCGGGTTCTCTTCCGGTGCACAAATTGGCATTCACGACGAGAAATATCTTG 22564
QY 241 gtaactgagatctcgtcgtgacacatctcaactacatcgtcgtcgtgacgacgtc 300
DB 22565 GCATTACGAGACTCTGCGCGCTCATTAATTTCTACACCTCGTGGTTATTCAGAGTTC 22624
QY 301 -----gtgaacatgttgcgcgcgccggggaatccccgcacatcatalcogaa 351
DB 22625 ACACATATGCGCATGCTGCTGCCGACAGGCGCAGTCCGCGCATGCTGGCT 22684
QY 352 gttgattcgtcgtatcgtacgctccgaatggatataccttcaagaagaagatcca 411
DB 22685 GTGAGCGCTGTGACTTGAAGTTCATCAGAGCGCTTGATCCGATCGGTTGAGAGCTCCG 22744
QY 412 tctaccggtatctgaataacacacacacacacacacacacacacacacacacacac 471
DB 22745 AGCAGACATATCTACAGTACACATCCGGTTCGACCCGCGCGCGGTGTTCCTC 22804
QY 472 tccacacagaagctcgggttaatttcgaacagctgattgtctgacttggagatcc 531
DB 22805 TCCACACAGATATCTACACCACTGCTGATGATGCTCAATCGGAGATTTCG 22864
QY 532 gacggattccacacgcaaatccgactcgtatcctggtacacacacacacacacac 591
DB 22865 GAAAGAGTTCG-----TCACCTCCCGTATCGTGGCGCTTCATCAGACATG 22915
592 ggtttgtaataagaaattcgacacacacacacacacacacacacacacacacacacac 651
DB 22916 GGGTTATGCTGGGATCATTTCTCCGATGATCAATCAACACACCGGCTGTGATGAGT 22975
QY 652 ccggtgtcgtcctcgacgcccgcgcggtgatatgaactgagtcgaacgatttacc 711
DB 22976 CCGATGGCATTTCTGCAAGCGCGCGCGCTGATGCAATTAATGCAACACCGCGCA 23035
712 gaccttcgacgacgcaaatctcgaccttgaactagcggcagcaagaacacacacacac 771
DB 2306 CAGATTTCAGCGCCACGATTTCCGTTTCGATGAGTGGCGGTACGAGAACATCCGAGAC 23095
QY 772 gacatgcccggcgctgactcggcaacatactgacacacacacacacacacacacac 831
DB 23096 GATATGCTGAGACTTACCTCGCGGATGTCGAGCATGCTCAACCGCGCGAGAGTA 23155
QY 832 caggcgacgacgac 881
DB 23156 AACGTGCGACGCTGCGGCGCTTCACGAGAGGTTGCGCCCTTCAACCT 23205

RESULT 14

MTCT338 29372 bp DNA BCT 17-JUN-1998
LOCUS Mycobacterium tuberculosis H37Rv complete genome; segment 127/162.
DEFINITION 274697 AL123456
ACCESSION 274697.1 GI:3261602
VERSION
KEYWORDS
SOURCE Mycobacterium tuberculosis.
ORGANISM Mycobacterium tuberculosis
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacteriales; Mycobacteriales;
Mycobacterium.

REFERENCE 1 (bases 1 to 29372)
AUTHORS Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C.,
Harris, D., Gordon, S.V., Eigmeier, K., Gas, S., Barry III, C.E.,
Tejeda, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T.,
Comnor, R., Davies, R., Devlin, K., Felwell, T., Gentles, S.,
Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J.,
Moule, S., Murphy, L., Oliver, S., Osborne, J., Quail, M.A.,
Rajandream, M.A., Rogers, J., Rutter, S., Seeger, K., Skelton, S.,
Squires, S., Squires, R., Sulston, J.E., Taylor, K., Whitehead, S. and
Barrell, B.G.

TITLE
JOURNAL
MEDLINE
REMARK
REFERENCE
AUTHORS
TITLE
JOURNAL

Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence
Nature 393 (6685), 537-544 (1998)
9825987
Erratum: [[published erratum appears in Nature 1998 Nov 12;396(6707):190]]
2 (bases 1 to 29372)
Parkhill,J.
Direct Submission
Submitted (11-JUN-1998) Submitted on behalf of the *Mycobacterium tuberculosis* sequencing and mapping teams, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA unite de Genetique Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk
On Jun 27, 1998 this sequence version replaced gi:2245330.

Notes:
Details of *M. tuberculosis* sequencing at the Sanger Centre are available on the World Wide Web.
(URL, <http://www.sanger.ac.uk/projects/M.tuberculosis/>) CDS have been renumbered from the original cosmid submissions but the old gene designations are in brackets after the new gene numbers.
Gene prediction was based on a Hidden Markov Model of TB genes implemented in TParse (Krogh) supplemented with visual inspection of positional base preference in codons, especially where there is an increase in the observed/expected third position G + C.
CAUTION: In some cases we may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, ggg, or tgg) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.

FEATURES
source
location/Qualifiers
1..29372
/organism="Mycobacterium tuberculosis"
/strain="H37Rv"
/db_xref="taxon:1773"
/clone="Y338"
complement(64..1821)
/gene="pkn1"
complement(64..1821)
/gene="pkn1"
/note="RV2914c, (MTCY338.02c), len: 585, probable serine/threonine-protein kinase similar to many in *M. tuberculosis* and other bacteria eg PKN1_MYXXA_P33973 serine/threonine-protein kinase pkn1, (693 aa) FASTA scores, opt: 286, z-score: 296.9, E(): 5.4e-10, (29.98 identity in 374 aa overlap); also similar to MTCY49.28 (34.7% identity in 479 aa overlap)"
/codon_start=1
/transl_table=11
/product="pkn1"
/protein_id="CA98986.1"
/db_xref="gi:3261603"
/translation="MALASGVTFAGYTVVRMLGCSAMGEVYLVQHPPGNOALKVLSPAMADDEFRRREFREVARLPHFHLLEHDEFGOLIMADYVDGIDATDHMA DRPAVLDFGEVGLAVITVAAGALDVAHGRVLPANVAVLTSOGADRIIADFGIAOSPSYAPRLSGADVDRADYALALTRIRHFACAPPYDRSHRTPLOPKISAR RPDARLDGVLAKLATAPADPFGSCREFADAMNEAGVAITADSSGVDSEVAAA GEEAVYVDYPAVGWPEAVDCKEFSRAPAPAPPPORGSMLQSAAGVLAARLDNEST ATKAPASPTRRRRRIIVGAAVILLAGLAVGIVIGRKTMTTATEAVARPTSGSAV SAPTTVAATVAVPLDGTYRIEIOHSKOTYDTPPODPDVTWMAFRPSGPECLA LATMIDNDHQAQPPYRPFIMOGEGOMKSRPETVOFCVGPNGSPSTQATQDLA LRPOQGLVGEVYTVVTSNECGOAGVIRIPANAASRGDLPRAVTVDPATIPDPTD TTSTATLPPTTATGPGR"

gene
CDS
complement(1865..2977)
/gene="RV2915c"
complement(1865..2977)
/gene="RV2915c"
/note="RV2915c, (MTCY338.03c), unknown, len: 370, similar to YUV3_YEAST_P40896 hypothetical 35.9 kd protein in *hbt-8* (331 aa) FASTA scores, opt: 228, z-score: 264.5, E(): 3.5e-08, (26.6% identity in 237 aa overlap). Also some similarity to AL021428|MTV030_18 *Mycobacterium*

gene
CDS
complement(3005..4582)
/gene="ffh"
/note="RV2916c, (MTCY338.04c), ffh, signal recognition particle protein, len: 525, similar to eg SR54_BACSU P37105 signal recognition particle protein (fifty-four homologue, (446 aa) FASTA scores, opt: 1451, z-score: 1409.6 E(): 0.5mlth-Waterman score: 1451; (51.5% identity in 435 aa overlap); contains P500017 ATP/GTP-binding site motif A"
/codon_start=1
/transl_table=11
/product="ffh"
/protein_id="CA98978.1"
/db_xref="gi:1405951"
/db_xref="SWISS-PROT:Q10963"
/translation="WFESLSDRLTALDGLRGKRLDADIDATRTRELRALLADVS LPVRAFTHRIKERAGAEVSSALNPAQOVKIVNEELISLIGETRELRAPKPPV VMLAGLQSGSKTTLAKGLARLRGSGHPLLVACDLPAAVNOLOVGEAEAPVFA PHGASPESGDPPVAVAAAGLAERAHFPPVVIYDTPRGAGIDELAAOAAIRDAI NPDEYLFVLDAMIGODAVTTAAAGCEGCFPGVALLTKDGGARGAALSVREVPVI LPASTGEKLEDFDVHNPMAASRIIGKMDVLSLIDQAKQVDDAQAEEAAKTAGEL TLEDFLEQMLAVRKMGPTIGNLLGPAQOKDALEVDQDLQDLVQAIINGMPQER ADPKTIINASRLRIANGSVTVSEVNOIVEFERKMMSSMLGGMIPGIRKSATP KSKGAKGSKGKSKKGTGRPTPPKVSFPFGVPMGLGLPGDLDSQMPKGDLEP PGLADPFDLSKLPFGPK"

misc_feature
gene
CDS
/note="RV2917, (MTCY338.05), unknown, len: 626; highly similar to *M. leprae* hypothetical protein MLCB250.18c (596 aa), fasta scores: opt: 3117 z-score: 3991.4 E(): 0, .79.88 identity in 584 aa overlap. Contains P500017 ATP/GTP-binding site motif A"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv2917"
/protein_id="CA98979.1"
/db_xref="gi:2245331"
/db_xref="SWISS-PROT:Q10966"
/translation="MRVRIVDADESTRKQDVCVPAPKSVMLHFTATSRFLGKIRANS VRSOGKVIQDPVSTFNPRLRGMQRALVLYLGTPQDPDLAVATPGSGKSPALRIA AELLRTHTVEQVTVVPTENLRKQVAHAAGHGLSDRKFANSNPOTSPEVIGVAVTY AOVASHPTLHRVTRT EARKTLVVFDEIHHGDAKTWGDAIRAFGDATRRLLLTGPF RSDSEIPFVSVOYPADGVLRQADHTYGYAALADGVVPAVFLFASQAQMRKSAGE EYEARLGEPLSAEQATRAMRTALDEPGEMVAIVTAAARRLRQLARHAPDAGCMITAS DRTTARAYARLLITMTMAEETPVYISDDSGSARSTIEFPOGTSRMILVAAWMSQGDVY RLNGVYATNASTPLFPQAIGRPVRSRRPSETASIFVSPVNLQTLASALEVQGNHY LGRPHRESAHDPDLSDPATVTRQTERGGAERKFTALGADAEIDVYTFDSSSFTATPTG SDEADYDGIPTGLDADQMRALLHNRQDQRLKRAVLODKGATQATPGASASVHGOLR DLRRRLHTLVIAHHRTKGPHGWIHDERRRCGPPIAAATRAQIKARIDALROLNRS"

misc_feature
4921..4944

```

/gene="Rv2917"
/Note="PS0017 ATP/GTP-binding site motif A"
complement(6551..8977)
/gene="glnD"
complement(6551..8977)
/gene="glnD"
/Note="Rv2918c, (MTCY338.07c), len: 808, glnD, Probable
uridylyltransferase, similar to eg glnD, ECOLI P27249
[protein-pil] (890 aa) FASTA scores, opt: 306, z-score:
339.0, E(): 2.5e-12, (27.7% identity in 857 aa overlap)."
/codon_start=1
/transl_table=11
/product="glnD"
/protein_id="CA98971.1"
/db_xref="GI:1405954"
/db_xref="SWISS-PROT:Q10961"
/db_xref="MEASPCASDLAVARRELLSGNHEDLPVGLRQTLWDLHESWL
/translation="MEASPCASDLAVARRELLSGNHEDLPVGLRQTLWDLHESWL
IKADIGIADASGFAIVGVGIGRELLPYSDLDLHLHDGPAIDLRVADRLWP
LMDANIRLHDSVTVSEALTIANSIDMAELGMEAHNIDGDOLEFALIDGVRQMN
GIRSMGEIVEMTVARMRCGRJAQRAEDPKIGRGGLRDVOLDLALAOQLDRGI
GHTDLPAGSLDGAIVRTLLDVRTLEHVSGRBHLAOFDEISALIGGDRDLART
LSAGRTIGYHAAGLRTANALPRGICSAIVRPRRRLDEGVYAGEIYARAE
PEHDPCLVLRVAASNDTGLPICAATLSRLAASVPLPTPMPEADLDLVLSAEP
TVATTEALDRGTGLMGRLLPEWEPIRLPPDVANHKVTVRHVETVHAAPLTVAR
PDLALGLLHDIGRGTDSVLAELVLPVCTRLGSPDVRTLSKLRHLLPLI
TATRLNDPKTIEAVSEALGDPOLLEVLHLSADSKATGDSVDMKASLVDLV
RCRMVWAGESLPOAERTPHYLSLAADHGVHEISPRDGERIDAVIVAPDERGLYSK
AAAVLANSLRVHSASVNVHOGVATIEFVYSPSPPAEIVROOFGALNDDVVL
GHLQKSDNASISVSARAGDVQAVPYTTPAPRLMLDTAPPAKLILEVRMDASG
LIALAGALEGAGAGIVMAKVNTEFSGTADVFCTVPAELDARAFAVEHILLEVLASV
DIVVDEPVD"
/complete(9035..9373)
/gene="glnB"
complement(9035..9373)

```

Query Match 34.7%; Score 313.4; DB 2; Length 29372;
 Best Local Similarity 60.6%; Pred. No. 8.5e-57;
 Matches 534; Conservative 0; Mismatches 341; Indels 6; Gaps 1;

```

QY 1 gattagaacagagactgagcgcgctgtgcataacccctgacgcgctgcagcgctgacg 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 22078 GACTACGAGATCCGACCCCAAGGAGATTGTGACGCTTGCTGCGACGCTTACAGT 22137
QY 61 cgaagcgtgaatgtgacacagagctgacgctgtgtgtccacaggtgacgcgctgtg 120
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 22138 CGTGCAATGCATCAATGCTGAAGAACAATCAATGAGGGGTACCGGAGATTCAGTGGC 22197
    121 atctctgtccgaaggaactcgatcgtcgcctatctcggcgctgtgcagggccgg 180
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
22198 GTTTTAGCCGCCACAGGACTGGAATATGCTCTGCAATCCCTGGCGGCACATTCAAGGCTGGA 22257
QY 181 cgaatcgcgctgcgctgtgttcacaaaggcgcggttcacgaatgagcttcgattcg 240
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 22258 TTTATGCGGCTTCGCTGTCACTCCACAGTATGGCAATTCACAGATGACCGCGTTTCTGCG 22317
QY 241 gtaactgagtgatcgtcgcgctgtgcacatctcactacatcgtcgcgtgcagcgatc 300
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 22318 GTGTTCAGAGATTCCAAAGCCGGTAAAGCATTCACGATTCGTCGCCGTAGGAGATGTA 22377
    301 gtgcacatgtgtgcgcgcgccggggaatccccgcacatcaatlcgaagtgtattg 360
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 22378 ACCAAATACCCACGACCCGACGACGCGGACCTGCCCCGCTCGATGTGAGTTGATCTG 22437
QY 361 ctgacatcgagcgctcgcaatgtgtataccttcaagaagaagatcatcatcaccg 420
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 22438 CTGTGATTGGACTGCGCGCACAGATGCCGGCTTTCTCTGTGACGACACACCGGGCGGCT 22497
QY 421 tatttgaatacacctcgcggttcaccccgacgcccgtgtgtgtgtatgctccatcag 480
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 22498 TATCTCCAAATACACGTCGCGATGACCGCGTACCGCGCGGAGATCATGTGTGCAACAG 22557
    481 aacgttcgggttaacttcgaacagctgtatgtctgtacttctcgcataccagcgagatt 540
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 22558 AATGTCATGCGCAATGTGACACAAAGTATGTACGCTATTTCGGCGATCCCGCAAGATT 22617
QY 541 ccaccgccaattccgactcgatctcctgctaccctctaccacgacatgggttgta 600
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 22618 CCGACCGGAGCTG-----TGCTGTCTGGCTGCTCTTTGTATCAAGATATGGGCTTAT 22671
QY 601 atagaatttgcacacacatcttggtgtgatacccgcggtgtctaccacgcccgtgtcg 660
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 22672 CTGCGAATTTGGCGACCGCGTGTGTGCGCCGACCGCGCGATGTTGATGACGCCAATGTCA 22731
    661 ttcttcagcgcccgccgcggtgtgatactgtatgctgcagcgatlttcacgcttcgt 720
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 22732 TTTTTCGCGCGCTCGCGCCCGCTGATGCAACTGCTTCCACAGCGGCGGTGCTTTCT 22791
QY 721 ggcagcagaatttgccttgaactgagcgacgaagaacacagcgagacatgccc 780
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 22792 GCGGACCGGAATTTGCTTTCGAGCTGTGCGTGGCGGAACATCTGACGAGACATGGCG 22851
QY 781 ggcgctgactcgcaacacatctgaccatctcagcggtagcgagcggtacagcgccg 840
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 22852 GGGCTGACCTGCGCGACGCGTGTGCGATGCTGATGCGAGTACGCAATTCATGTGCA 22911
QY 841 agatcaagcgcttcgcgacgcgcttgcgtctcaatc 881
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 22912 ACCGTGCGCGGTTTCATCGAGCGGTTGCGCGCGTACATCT 22952

```

RESULT 15

LOCUS MLCB12/c

DEFINITION Mycobacterium leprae cosmid B12.

ACCESSION AL035480.1 GI:4455662

VERSION AL035480.1 GI:4455662

KEYWORDS 19 kD antigen; 19 kD lipoprotein; acyl-CoA dehydrogenase; acyl-CoA synthase; pksB; polypeptide synthase; pseudogene; RLEP, thioesterase.

SOURCE Mycobacterium leprae.

ORGANISM Mycobacterium leprae

Bacteria: Firmicutes: Actinobacteria: Actinobacteridae: Actinomycetales: Corynebacterineae: Mycobacteriaceae; Mycobacterium.

REFERENCE 1 (bases 1 to 34316) Elgmeyer, K., Honore, N., Woods, S.A., Caudron, B. and Cole, S.T. Use of an ordered cosmid library to deduce the genomic organization of Mycobacterium leprae

JOURNAL Mol. Microbiol. 7 (2), 197-206 (1993)

MEDLINE 9318700

REFERENCE 2 (bases 1 to 34316) Murphy, L. and Harris, D. Unpublished

AUTHORS JOURNAL 3 (bases 1 to 34316) James, K.D., Parkhill, J., Barrell, B.G. and Randal, M.A. Direct Submission

REFERENCE Submitted (15-FEB-1999) Mycobacterium leprae sequencing project, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Dr. Stewart T. Cole, [3] Unite de Genetique Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France Requests for cosmids should be sent to Karin Elgmeyer (kelepasteur.fr)

NOTES:

The Sanger Centre is funded to complete the sequence of M. leprae by the Heiser Program for Research in Leprosy and Tuberculosis of The New York Community Trust.

Work in Paris is supported by the Heiser Trust, the Association Francaise Raoul Follereau and the Groupement de Recherches et des Etudes des Genomes (GIP-GREG).

Details of M. leprae sequencing at the Sanger Centre are available on the World Wide Web.

(URL, <http://www.sanger.ac.uk/projects/>)

CDS are numbered using the following system eg MLCB33.01c. ML (M. leprae), cb33 (cosmid name), .01 (first CDS), c (complementary strand).

The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous. The length in codons is given for each CDS.

Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database. The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon. All CDS over 100 codons have been analysed. Gene prediction is based on positional base preference in codons especially where there is an increase in the observed/expected third position G + C. CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, or ttg) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.

IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions.

```

FEATURES
source
    1..34316
        /organism="Mycobacterium leprae"
        /db_xref="taxon:1769"
        /clone="cosmid B12"
        complement(1..9070)
        /note="overlap with EMBL:U00023 cosmid L518 from 9070 to 1. There are 3 conflicts between this sequence and ours. In each case our sequence has been checked and is thought to be correct"
        complement(52..4674)
        /gene="MUCB12.01c"
        complement(52..4674)
        /gene="MUCB12.01c"
        complement(52..4674)
        /note="MUCB12.01c, probable polyketide synthase, Jan: 1540 aa; similar to ERYL_SACBR (EMBL:M63676) Saccharopolyspora erythraea erythronolide synthase (3491 aa), fasta scores: opt: 2230 z-score: 2283.7 E(): 0, 36.6% identity in 1550 aa overlap. N-terminus similar to the N-terminus of MCAS_MYCB0 (EMBL:M95808) Mycobacterium bovis mycoerotic acid synthase (2110 aa) (42.2% identity in 879 aa overlap). Also similar to many mycobacterial putative polyketide synthases e.g. ppsb, Rv2932, (MTV011.01-MTCY338.21) M.tuberculosis putative polyketide synthase involved in phenolphthiocerol synthesis (1538 aa) (76.3% identity in 1561 aa overlap). Annotated as ORF TR:Q49932, designated pksc in M.leprae cosmid EMBL:U00023. Contains Pfam match to entry PF00550 pp-binding, Phosphopantetheine attachment site, score 42.90, E-value 3e-10, Pfam match to entry PF00698 Acyl-transf. Acyltransferase domain, score 467.00, E-value 1.6e-136, Pfam match to entry PF00109 ketoacyl-synt, Beta-ketoacyl synthase, score 694.00, E-value 7.4e-205, Pfam match to entry PF00501 AMP-binding, AMP-binding enzyme, score -162.30, E-value 4.9e-06. Contains PS00606 Beta-ketoacyl synthases active site"
        /codon_start=1
        /transl_table=1
        /product="putative polyketide synthase"
        /protein_id="CAB36627.1"
        /db_xref="GI:4455663"
        /translation="MRTAFSRISGMTTQORAALEETFKLSRIAVERTAVVIGICRP
        PGDVTGPDSPFDLLIDGRNATSRVPADWDADATYDPDLTPGKMTTKWGFVSIDAG
        FDAAFGITTREAAAMDPORIILEVAWEALNAGIPDSIGNSRTGVMICVINEYO
        SMLASLENDVAYSGETGNASITVGRISYLLGLRGPVAVDTACSSSLVAHLACOSL
        RLRTEDVLAVGVSITLRFETQIAISAGLISPHGRCAAFPAADAGFEVRSAGAVVYL
        KRLIDAVDGLVLAVRGSVAVNODGSRNGVTAPNTAACDVITDALLSSSVAPESVY
        VYESHGCTVLDGPIEPALATATYGRCSACATGAVKNTCHILAAKATIAFITVLA
        VORQIPTNLHFSOMNPAIDASTREFPTENSSPITDGGARRAVSVSSFGUGATN
        AHVIEGDELTPEVTESSNTAVSTLVVGTASRVAAAGMLADWEGPAEVALAD
        VAHTLNHRSHRAKFGTVVARDRIQVAAGLALAKOAPGVVCOQDTPSSGIVTVY
        SGRGSOWAGMRQLADPAFTAAVAVLEPVAAGSLVDVANGKEIAGIEIDQL
        GLIGMQLTLELMRSYGVOPDLVIGHSIGEEVAAVAGALTAAGLRVATARSRLMAP
        LSGGGMALLEDAVEETALLVDVYSVYTLAIYNSPRTVINGPTQEDIDELDRVRAQ

```

```

misc_feature
    347348
        /note="conflict: A is present in EMBL:U00023"
        1981..3367
        /note="1387 bp direct repeat"
        complement(2050..2991)
        /gene="MUCB12.01c"
        /note="Pfam match to entry PF00698 Acyl-transf. Acyltransferase domain, score 467.00, E-value 1.6e-136"
        complement(3298..4572)
        /gene="MUCB12.01c"
        /note="Pfam match to entry PF00109 ketoacyl-synt, Beta-keto acyl synthase, score 694.00, E-value 7.4e-205"
        complement(3886..4965)
        /note="Pfam match to entry PF00501 AMP-binding, AMP-binding enzyme, score -162.30, E-value 4.9e-06"
        complement(4042..4092)
        /gene="MUCB12.01c"
        /note="PS00606 Beta-ketoacyl synthases active site"
        complement(4675..10290)
        /gene="MUCB12.02c"
        complement(4675..10290)
        /gene="MUCB12.02c"
        /note="MUCB12.02c, probable polyketide synthase, Jan: 1871aa; similar to MSAS_PEN2A (EMBL:P2367) Penicillium patulum 6-methylsalicylic acid synthase (1774 aa), fasta scores: opt: 2727 z-score: 2873.8 E(): 0, 33.8% identity in 1823 aa overlap. Also highly similar to many mycobacterial putative polyketide synthases and equivalent to TR:EI299930 (EMBL:Z7469.7) M.tuberculosis ppsB, Rv2931 (MTCY338.20) (1876 aa) (76.1% identity in 1888 aa overlap). C-terminus annotated as ORF TR:Q49931, designated pkxb in M.leprae cosmid EMBL:U00023. Contains two Pfam matches to entry PF00550 pp-binding, Phosphopantetheine attachment site, score 77.40, E-value 8.8e-20 and score 61.30, E-value 2.5e-15, Pfam match to entry PF00698 Acyl-transf. Acyltransferase domain, score 467.00, E-value 1.6e-136, Pfam match to entry PF00109 ketoacyl-synt, Beta-keto acyl synthase, score 626.50, E-value 1.5e-184. Contains PS00142 Phosphopantetheine attachment site, PS00606 Beta-ketoacyl synthases active site and probable coiled-coil from 1748 to 1781 (34 residues)"
        /codon_start=1
        /transl_table=1
        /product="putative polyketide synthase"
        /protein_id="CAB36628.1"
        /db_xref="GI:4455664"
        /db_xref="SPTREMBL:Q92566"
        /translation="MTASIGSEADLCHLVNVTNIGCTPDEVDLNLSTDLGMSNR
        DAVVLSGETLIDLLKRTVSPIDFWEPHTINAALVLAIPAPDEESAPRRVQALAE
        PIAVIGMCRPGGISGPEALAMOLFCLRRSSIGGVNEMWELTFDGSSEVALARTT
        RMGSELDIDAFDEPFIEPSDEADKDPQORLLPEVMEWLEHAGLSNLSRSQGF
        VFAGSCISEYGAINSTDSQVDSNGMSGASITANRLSYTLDRGSAVAVDTCCSS
        LVAHLHACQSLRTADANLAIAGVNLILSPAVFPGPDVYCALSTPGCCRAFDAAADG

```

VREGAGVVLKRLTDAQRDQRYLAIIRGSAVNQDRSNGLMAPNPAQAVALTAY
ANAMPPTLIDYEHTGTGLDPIEARLGLVLRGRBEDPILIGVKTMLGLE
AAAGIAGFETVLAVOIGOIPNORESPNLIHPAELRMKVVDKHEMPDITGLRRA
AVSFGEGGNNAHVILEOGDELPVTECSNNTVSTLVYTKASVPAAMAGLADW
EGSGAVALADVAHNTLNHRSRAKRGIVARIGVLRGLAAGKAQPGVGOO
GTPGSGTVFVYSGRSQWAGMKOLLADEPAFAAAELEPPVFIHAGSLHVLNAG
KELVIGIEQIGLIGMQLTLELMRSYGVDPDVLVIGHSIGEVAAVAVAGLTAEGLR
VYATRSRLMAPLSGGGMALLELDVATEVLEIVDSOVLAIYNSPROTVIAGPTEOI
DELIDVRAONRFRASRVNIEVAPHNMDALOPMRSELDVAPRPTTPIILSTYAD
LGSPVFDQAHMATNMNRNPVHFQQAIMTGMTDHRFLEISAPHLTLQAITDLHGTRC

Query Match 32.28; Score 291; DB 2; Length 34316;
Best Local Similarity 59.08; Pred. No. 4.3e-52;

Matches 520; Conservative 0; Mismatches 355; Indels 6; Gaps 1;

QY 1 gattacgaacagagactgagacgagctgtgcataaccctgacgagctgcagctgtatcgg 60
11954 GACTACGGGTGGACCGGAAAGGTTTGTGAGAGCTTGACGTGGTCGCAAGTTTATATG 11895
61 cgaacgctgaatgtgacagagagctgagccgtgtgtgtccacggtgacccgctgtg 120
11894 CGTTTCATGCGCTTGTCTGACGAACTCAAAATGTGTGTGGTGGCCGGTGATCGAGTTGCA 11835
QY 121 atctctgtctcgcgaagagactcgagtaactgtcgtccatctcggcgctgcagccgg 180
11834 ATTTTGGCGGCACAGGAGCTGGAATATATTATCGCTTTAGGGCTCTTCAAGCCGGA 11775
QY 181 cgcacgcgctgcgagcttcacaaagcggttaccgaatgacgttccgactcg 240
11774 TTTATCGCGGTACTCTTTTGGCCACCGCAGATTCGATTCACGACGAAACGGCTTTCGGCG 11715
QY 241 gtaactgagtaattcgttcgagtggtccattctcaatacgtctgcgtgagacgctc 300
11714 GTGTTACGGGATCTCTCAACGATCGTATTCTTACACTTCATTGTGTAGTAGACGTG 11655
QY 301 gtgcacaatgttcgagcggtcccgagggaatcccgccatcaatatacgaagtgtatg 360
11654 GCGAAATACGCATGTGAACAAGATGGGTGAACGCTCGTGTGCATTGAGATCGATCTG 11595
QY 361 ctgcactgagagctccggaatggtataacttcaagaagagatcatcattacccg 420
11594 CTGACTTGAACCGGTACCTCCGCTTCCAAAAGTCCCCAGCCTTTACTAGCCCGCT 11535
QY 421 tattgcaatacacctccggtgtccaccgacgcccgtgcgtgtgtatgtccatcag 480
11534 TACCTCAATATACATCCGGGTGACACGTACGCCACCGGTATTATGTGTGCATATAG 11475
481 aaggttcggttaatttcgaacagctgagtctcgtacttgcggtatccgaacgagatt 540
11474 AACGTTATCGCTTAACGTGACGCAAGCATGACGGCTATTTCGGCATVGGCGCAAAAAC 11415
QY 541 ccacgcgcaaatccgactcgtatccttgtaaccttaccacgacatggtttgtgta 600
11414 CTG-----AACACGACCGGTGTGCTATGCTTCCCTTGTTCACGACATGGGCTTGATT 11361
QY 601 ataggaaatttgacccaattctggtgtgatacccggtgtcctcaaccagccggtgtcg 660
11360 CTTGGGGTCTGTGCTCAATGTGCGGGGCGAAGCTCGTGTGTGAGCCCTATATGCA 11301
QY 661 ttctctcagagccggtccggtgtgatacttgatgtgcagcgatlttcaagcctttcg 720
11300 TTTTTCGGGGCCCCCGCTGTGTGATGCAACTTCTTGCCACTACCAAGGGGTGTTTCT 11241
QY 721 gcagacacgaatttcgcttgaactagcgacgaagaacacgacgacgacatgagcc 780
11240 GCGGCAACCGAATTTTGCTTTGAGCTGGCGGTACGCAAGACGTCCGACGACATGAAAC 11181
QY 781 gggcgctgacctcgcaacatctgacatcctcctcagcgttagcagcgggtacagccgg 840
11180 GGACTCGACTCGGGGATGTGGCGGATCATGACGGCGACGAGTCAATATCGCA 11121
QY 841 acgataagcgttcgacgacgcttgcctcaatct 881

Db 11120 ACGGTGAACGTTTCACCGAGCGCTTCGCTCGTTCAATCT 11080

Search completed: January 12, 2001, 20:06:54
Job time: 26588 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 5, 2001, 14:30:37 ; Search time 115.27 Seconds
(without alignments)
305.044 Million cell updates/sec

Title: US-09-461-774-6

Perfect score: 1540
Sequence: 1 DYEQDMGVATITLRSQLYR.....TIKRFADRFARFNQERVKA 300

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: SP archaea:*
2: SP bacteria:*
3: SP fungi:*
4: SP human:*
5: SP_invertebrate:*
6: SP_mammal:*
7: SP_mhc:*
8: SP_organelle:*
9: SP_phage:*
10: SP_plant:*
11: SP_rudent:*
12: SP_virus:*
13: SP_vertebrate:*
14: SP_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1519	98.6	580	2	P96290 mycobacteri
2	1515	96.4	582	2	P71493 mycobacteri
3	954.5	62.0	584	2	O53903 mycobacteri
4	948	61.6	584	2	O07797 mycobacteri
5	906.5	58.9	578	2	O50441 mycobacteri
6	858	55.7	583	2	O925K5 mycobacteri
7	823	53.4	619	2	P95141 mycobacteri
8	821	53.3	411	2	O50466 mycobacteri
9	604.5	39.3	585	2	P95213 mycobacteri
10	484	31.4	637	2	O53580 mycobacteri
11	426	27.7	620	2	P95288 mycobacteri
12	421	27.3	2297	2	O54155 streptomyc
13	386.5	25.1	1770	2	O50857 myxococcus
14	310	20.1	8243	5	O86554 cryptospori
15	290	18.8	110	2	O86982 mycobacteri
16	271	17.6	220	2	P72116 pseudomonas
17	257.5	16.7	562	2	P71605 mycobacteri
18	239	15.5	587	2	O929G2 burkholderi
19	233	15.1	649	10	Q9XHW1 oryza sativ

20	232	15.1	544	2	O05598 mycobacteri
21	223	14.5	3971	2	O9R9J1 bacillus su
22	199.5	13.0	485	2	O910A2 streptomyc
23	195.5	12.7	511	2	O9K2C1 streptomyc
24	195	12.7	1400	2	O91C00 mycobacteri
25	193.5	12.6	4450	2	O44928 bacillus br
26	193.5	12.6	7463	2	O924X6 streptomyc
27	190	12.3	4976	2	O87314 mycobacteri
28	187	12.1	2631	2	O9RNB5 microcystis
29	185	12.0	3643	2	O9RK14 streptomyc
30	183	11.9	3415	2	O9XCF3 mycobacteri
31	182.5	11.9	525	2	O07169 mycobacteri
32	181.5	11.8	584	2	O9RTR4 oryza delinococcus
33	181.5	11.8	1575	2	P94904 lysobacter.
34	181	11.8	1876	2	O9R686 bacillus su
35	179.5	11.7	3317	2	O9RAH2 nostoc sp.
36	177.5	11.5	8563	2	O54297 streptomyc
37	177	11.5	3670	2	O924X5 streptomyc
38	176.5	11.5	9376	2	O85168 pseudomonas
39	173	11.2	522	2	O9K3W1 streptomyc
40	173	11.2	7576	2	O92GA4 streptomyc
41	172.5	11.2	476	2	O50017 mycobacteri
42	172.5	11.2	2591	2	O54959 streptomyc
43	172	11.2	550	5	O27758 photinus py
44	172	11.2	1394	5	O01787 caenorhabdi
45	171.5	11.1	2611	2	O6487 streptomyc

ALIGNMENTS

RESULT 1
ID P96290 PRELIMINARY; PRT; 580 AA.
AC P96290;
DC 01-MAY-1997 (TREMBLrel. 03, Created)
DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE HYPOTHETICAL 62.6 KDA PROTEIN.
GN RV2941 OR MTCY24G1.08C.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxId=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D., Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F., Badcock K., Basham D., Brown D., Chillingworth T., Connor R., Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holtroyd S., Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L., Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J., Rutter S., Seeger K., Skelton S., Squares S., Sulton J.E., Taylor K., Whitehead S., Barrell B.G.;
RA *Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.*;
RL Nature 393:537-544 (1998).
DR EMBL: 283858; CAB06100.1; -
DR TUBERCULIST; RV2941; -
DR INTERPRO; IPR000873; -
DR INTERPRO; IPR002048; -
DR PFAM; PF00501; AMP-binding; 1.
DR PROSITE; PS00018; EF_HAND; UNKNOWN_1.
KW Hypothetical protein.
SQ SEQUENCE 580 AA; 62640 MW; 60842A083746AFD3 CRC64;

Query Match 98.6%; Score 1519; DB 2; Length 580;
Best Local Similarity 99.3%; Pred. No. 4,7e-119;
Matches 296; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```

QY 1 DYEDMDGVAITLTRSOLYRRTLNVAOELSRCSGTGRVVISAPQGLEVVAYLGAALQAG 60
DB 27 DYEDMDGVAITLTRSOLYRRTLNVAOELSRCSGTGRVVISAPQGLEVVAYLGAALQAG 86
QY 61 RIVAPLSVPGGVYTDERSVLSDSPPVAILTTSSAVDDVYQVHARRPPGSPSIIIVDL 120
DB 87 RIVAPLSVPGGVYTDERSVLSDSPPVAILTTSSAVDDVYQVHARRPPGSPSIIIVDL 146
QY 121 LDLDAPNGYTFKDEYVSTAYLOYTSGSTRTPAGVYVSHQNVNRFQMLMSGYFADTDGI 180
DB 147 LDLDAPNGYTFKDEYVSTAYLOYTSGSTRTPAGVYVSHQNVNRFQMLMSGYFADTDGI 206
QY 181 PPNSALVSLVPEYHDMGLVIGICAPILGGYPAVLTPSVFLOQPARMHLMSDFHAFS 240
DB 207 PPNSALVSLVPEYHDMGLVIGICAPILGGYPAVLTPSVFLOQPARMHLMSDFHAFS 266
QY 241 AARNFAFELAAARTTDDDMAGRDGLNLTLLSGSERVOAATIKRFADRFARFNLOERV 298
DB 267 AARNFAFELAAARTTDDDMAGRDGLNLTLLSGSERVOAATIKRFADRFARFNLOERV 324

RESULT 2
P71495 PRELIMINARY: PRT: 582 AA.
ID P71495;
AC P71495;
DT 01-FEB-1997 (TREMBLrel. 02, Created)
DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE ACYL-COA SYNTHASE.
OS Mycobacterium bovis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1765;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BG;
RX MEDLINE=97252490; PubMed=9098059;
RA Fitzmaurice A.M., Kolattukudy P.E.;
RT "Open reading frame 3, which is adjacent to the mycocerosic acid
RT synthase gene, is expressed as an acyl coenzyme A synthase in
RT Mycobacterium bovis BCG."
RL J. Bacteriol. 179:2608-2615(1997).
DR EMBL: 075685; AAB52538.1; -.
DR INTERPRO: IPR000873; -.
DR PFAM: PF00501; AMP-binding; 1.
SQ SEQUENCE 582 AA: 62930 MW: 44877609A8EEA552 CRC64;

```

Query Match

Best Local Similarity 98.4%; Score 1515; DB 2; Length 582;
Matches 295; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

```

QY 1 DYEDMDGVAITLTRSOLYRRTLNVAOELSRCSGTGRVVISAPQGLEVVAYLGAALQAG 60
DB 29 DYEDMDGVAITLTRSOLYRRTLNVAOELSRCSGTGRVVISAPQGLEVVAYLGAALQAG 88
QY 61 RIVAPLSVPGGVYTDERSVLSDSPPVAILTTSSAVDDVYQVHARRPPGSPSIIIVDL 120
DB 89 RIVAPLSVPGGVYTDERSVLSDSPPVAILTTSSAVDDVYQVHARRPPGSPSIIIVDL 148
QY 121 LDLDAPNGYTFKDEYVSTAYLOYTSGSTRTPAGVYVSHQNVNRFQMLMSGYFADTDGI 180
DB 149 LDLDAPNGYTFKDEYVSTAYLOYTSGSTRTPAGVYVSHQNVNRFQMLMSGYFADTDGI 208
QY 181 PPNSALVSLVPEYHDMGLVIGICAPILGGYPAVLTPSVFLOQPARMHLMSDFHAFS 240
DB 209 PPNSALVSLVPEYHDMGLVIGICAPILGGYPAVLTPSVFLOQPARMHLMSDFHAFS 268
QY 241 AARNFAFELAAARTTDDDMAGRDGLNLTLLSGSERVOAATIKRFADRFARFNLOERV 298
DB 269 AARNFAFELAAARTTDDDMAGRDGLNLTLLSGSERVOAATIKRFADRFARFNLOERV 326

```

```

RESULT 3
O53903 PRELIMINARY: PRT: 584 AA.
ID O53903;
AC O53903;
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE HYPOTHEICAL 62.8 KDA PROTEIN.
GN RV1529 OR MT0035.03.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squires R., Sulston J.E.,
RA Taylor K., Whitehead S., Barrett B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
DR EMBL: AL022000; CA017589.1; -.
DR TUBERCULIST; RV1529; -.
DR INTERPRO: IPR000873; -.
DR PFAM: PF00501; AMP-binding; 1.
KW Hypothetical protein.
SQ SEQUENCE 584 AA: 62829 MW: CC09FDEA5B0ABD5 CRC64;

```

Query Match

Best Local Similarity 62.0%; Score 954.5; DB 2; Length 584;
Matches 190; Conservative 35; Mismatches 63; Indels 15; Gaps 2;

```

QY 1 DYEDMDGVAITLTRSOLYRRTLNVAOELSRCSGTGRVVISAPQGLEVVAYLGAALQAG 60
DB 27 DYEDMDGVAETLTRSOLYRRLNVAERLHVGATGDRVILAPQGLEVVAYLGAALQAG 86
QY 61 RIVAPLSVPGGVYTDERSVLSDSPPVAILTTSSAVDDVYQVHARRPPGSPSIIIVDL 120
DB 87 RIVAPLPVPHAGAHDEITVLSDTSPVAILTTSGANDVYQVHARRPPGSPSIIIVDL 146
QY 121 LDLDAPNGYTFKDEYVSTAYLOYTSGSTRTPAGVYVSHQNVNRFQMLMSGYFADTDGI 180
DB 147 LDLDAPNGYTFKDEYVSTAYLOYTSGSTRTPAGVYVSHQNVNRFQMLMSGYFADTDGI 201
QY 171 SGYFADTDGIIPPNLSALVSLVPEYHDMGLVIGICAPILGGYPAVLTPSVFLOQPARMHL 230
DB 202 ADFEAPGCGVPPDLTVVSWLPYHDMGLLGAIMPLIAGVPTVLPVGLRPARMIQ 261
QY 231 LMSDFHAFSAARNFAFELAAARTTDDDMAGRDGLNLTLLSGSERVOAATIKRFADRF 290
DB 262 LMRNGRTISAGPFAFELAVRKISDDMDGLDLAGVHTLLNGSERVHPATLTKRFAERFG 321
QY 291 RFN 293
DB 322 RFN 324

RESULT 4
O07797 PRELIMINARY: PRT: 584 AA.
ID O07797;
AC O07797;
DT 01-JUL-1997 (TREMBLrel. 04, Created)
DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE POSSIBLE POLYPEPTIDE SYNTHASE.

```


OS Mycobacterium tuberculosis.
 CC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 CC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OK NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-H37RV;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Fellwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,
 RA Taylor K., Whitehead S., Barrell B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence."
 RL Nature 393:537-544(1998).
 RN [2]
 NC SEQUENCE FROM N.A.
 MC STRAIN-H37RV;
 RA Parkhill J.;
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: Z84725; CAB06604.1; -
 DR TUBERCULIST; Rv0404; -
 DR INTERPRO; IPR000873; -
 DR PFAM; PF00501; AMP-binding; 1.
 SQ SEQUENCE 585 AA; 63843 MW; B16A77089F760F18 CRC64;

Query Match 39.3%; Score 604.5; DB 2; Length 585;
 Best Local Similarity 44.3%; Pred. No. 1.9e-42;
 Matches 133; Conservative 47; Mismatches 112; Indels 7; Gaps 3;

QY 1 DYEODMDGVAITLRSQLYRRLTNAOELSRGSGTGDR--VVISAPGLEYVAVYLGAL 57
 DB 24 DYDTGTGQIDMTWSQSLKSRVTAVSALISYGRHADRRRTAISAPGLODYVAFLCAL 83
 QY 58 QAGRIAPVLPVPGGVTDERSDVSLSPPAIIITSSAVDDVVOHVARPESPSITE 117
 DB 84 CAGMPVLPPEPLGSLRKRTGLAVLDCAADVLTTSQETRVATITAGCASVTTPYA 143
 QY 118 VDLDDLPANGTTFKEDYPR--TAYLOYTSGSTRTPAGVNSHONVRNFEQLMSGTFA 175
 DB 144 LD-TLDEPSSGDCDLDSQLSDWSSYLQYTSSTANPRGVLSMRNVTENVQIIRNYFR 201
 QY 176 DTDGIPPPNSALVSWLPFYHDMGLVIGCAPILGGYPAVLTPSPFLOPARMHLMASD 235
 DB 202 HEGGAPRLPSSVSWLPLXHDGMLAVGLFPLFVGCPYLITSPEAFIRKPAKMQLLAKH 261
 QY 236 PAFSAAPNFAFELAAARTTDDMAGRDGLNLTLLSGSERVAATIKRFADRFARFNL 294
 DB 262 QAFSAAPNFAFELAAVAKTSEEDMAGLDLGHVNTIINGAEQVQPTIRKELRRFPYNL 320

RESULT 10
 ID 053580 PRELIMINARY; PRT; 637 AA.
 AC 053580;
 DT 01-JUN-1998 (TREMBLrel. 06, Created)
 DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
 DE 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
 DE PUTATIVE POLYKETIDE SYNTHASE.
 GN RV3801C OR MTW026.06C.
 OS Mycobacterium tuberculosis.
 CC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 CC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OK NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-H37RV;
 RX MEDLINE=98295987; PubMed=9634230;

RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Fellwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,
 RA Taylor K., Whitehead S., Barrell B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence."
 RL Nature 393:537-544(1998).
 DR EMBL: AL022076; CAAL7865.1; -
 DR TUBERCULIST; Rv3801c; -
 DR INTERPRO; IPR000873; -
 DR PFAM; PF00501; AMP-binding; 1.
 SQ SEQUENCE 637 AA; 69231 MW; 0A3A86CED9AE0EDC CRC64;

Query Match 31.4%; Score 484; DB 2; Length 637;
 Best Local Similarity 37.8%; Pred. No. 2.7e-32;
 Matches 115; Conservative 51; Mismatches 120; Indels 18; Gaps 5;

QY 1 DYEODMDGVAITLRSQLYRRLTNAOELSRGSGTGDRVISAPOGLEYVAVYLGALOG 60
 DB 50 DETEREDGVARDLMSDSARRNARGARLQOYTOPGDRVAILCPQNDLILSFFGALISG 109
 QY 61 RIAPVLPVPGGVTDERSDVSLSPPAIIITSSAVDDVVOHVARPESPSITEVDL 120
 DB 110 RIAPVLPFAEPAGHVGRIHNAVLDCAAPSYILTTTDSAGCVKRFIARSAKEPRVIAVDA 169
 QY 121 LDLDAPNGTTFKEDYPRATAYLOYTSGSTRTPAGVNSHONVRNFEQLMS--GYFADT 177
 DB 170 VPEEAATWQPEANEVEYVAYLOYTSGSTRIPSGVQITHLNPTNVVOVLANLEQEGDR 229
 QY 178 DGIPPNSALVSWLPFYHDMGLVIGCAPILGGYPAVLTPSPFLOPARMHLMASD 233
 DB 230 G-----VSMPLPFHDMGLITVLLASVL-GHFTFTPTAFVRRGRVRELAKRPG 279
 QY 234 SDFHAFSAAPNFAFELAA-RRTTDDMAGRDGLNLTLLSGSERVAATIKRFADRFARF 292
 DB 280 ETGTFSAAPNFAFHAAVRGVRDEPDLDSNVKGLINGSEPVSPASMRKFEAFAPY 339
 QY 293 NQGE 296
 DB 340 GLKO 343

RESULT 11
 ID P95288 PRELIMINARY; PRT; 620 AA.
 AC P95288;
 DT 01-MAY-1997 (TREMBLrel. 03, Created)
 DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
 DE 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
 DE HYPOTHETICAL 66.3 KDA PROTEIN.
 GN RV1925 OR MTCT09P9.39C.
 OS Mycobacterium tuberculosis.
 CC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 CC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OK NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-H37RV;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Fellwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,
 RA Taylor K., Whitehead S., Barrell B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the

RT complete genome sequence."
 RL Nature 393:537-544 (1998).
 DR EMBL: Z84498; CAB06501.1; -.
 DR HSSP: P08659; ILIC1.
 DR TUBERCULIST: RV1925; -.
 DR INTERPRO: IPR000873; -.
 DR PFAM: PF00501; AMP-binding; 1.
 DR KW Hypothetical protein.
 SQ SEQUENCE 620 AA; 66327 MW; 6858376B5F25BBC6 CRC64;

Query Match 27.7%; Score 426; DB 2; Length 620;
 Best Local Similarity 35.9%; Pred. No. 1.9e-27;
 Matches 108; Conservative 55; Mismatches 120; Indels 18; Gaps 8;

DB 1 DYQDDMGVAITLTSRQLRTLNVAOELSRGSGTGRVVISAPQGLEVVAYIGALQAG 60
 57 DHARSAAGCALLETWTQFGRLAIGAHOVRFAGPGDRAIILAPQIDYCGFYAAIKAG 116
 61 RIAVPLSVPGGVTDERSDVSLSDFVALITTSAAVDVYQVHARRPGSPSIIENVL 120
 117 TVAVPLFAPRLPGHAEHLDLTALDSEPAVILTTAAKNAVEGFLNNVRLRKPTVLIDQ 176
 DB 121 LDDAENGITFKDEYPTAYLQYTSSTRTPGVVMASHQNVNFEQLMSGTFADTGT 180
 177 IPREGELFVPEMDIDVAHSLQYTSSTRTPGVVEITHRAVGTNVLQMLST----DL 231
 QY 181 PPNNSALVSLPFYHMGVL-VIGICAPILGGYPAVLTPSVFLQRPARRMHLASDF--- 236
 DB 232 LNNTHGVSWLPLYHDMGLSMICFPA-VYGGH-STLMSPTAFYRRLRMQALSEGSRGT 289
 QY 237 HAFSAAPNFAFELAAARR--TTDDMAGRDGLNLTILSGSERVOATIKRFADRFARN 293
 290 RVTATAPNFAVEMAAQRGLPAQGDV--DLSNVLLIT-GSEPVSIDAVYTFKAFAPYG 345
 QY 294 L 294
 DB 346 L 346

RESULT 12
 OS4155 PRELIMINARY; PRT; 2297 AA.
 AC 054155;
 DT 01-JUN-1998 (TREMBLrel. 06, Created)
 DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
 DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
 POLYKETIDE SYNTHASE.
 SC3F7.12.

Streptomyces coelicolor.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
 NCBI_TaxID=1902;

RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RA Oliver K., Harris D.;
 RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RA Parthill J., Barrell B.G., Rajandream M.A.;
 RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RX MEDLINE-97000351; PubMed-8843436;
 RA Redenbach M., Kieser H.M., Denapalte D., Eichner A., Cullum J.,
 RA Kinashi H., Hopwood D.A.;
 RT "A set of ordered cosmids and a detailed genetic and physical map for
 the 8 Mb Streptomyces coelicolor A3(2) chromosome.",
 RT Mol. Microbiol. 21:77-96(1996).
 DR EMBL: AL021409; CAA16183.1; -.

DR HSSP: P25715; 1MLA.
 DR INTERPRO: IPR000255; -.
 DR INTERPRO: IPR000794; -.
 DR INTERPRO: IPR000873; -.
 DR INTERPRO: IPR001227; -.
 DR INTERPRO: IPR001917; -.
 DR PFAM: PF00109; ketoacyl-synt; 1.
 DR PFAM: PF00222; aminotran-2; 1.
 DR PFAM: PF00501; AMP-binding; 1.
 DR PFAM: PF00550; pp-binding; 1.
 DR PFAM: PF00698; Acyl transf; 1.
 DR PRINTS: PR00154; AMPBINDING.
 DR PROSITE: PS00399; AA-TRANSFER_CLASS_2; UNKNOWN_1.
 DR PROSITE: PS00606; B_KETOACYL_SYNTHASE; 1.
 DR PROSITE: PS50075; ACP_DOMAIN; 2.
 DR KW transiferase.
 SQ SEQUENCE 2297 AA; 241989 MW; C67B58461535EEA6 CRC64;

Query Match 27.3%; Score 421; DB 2; Length 2297;
 Best Local Similarity 36.4%; Pred. No. 3.2e-26;
 Matches 107; Conservative 38; Mismatches 133; Indels 16; Gaps 4;

QY 5 DMDGVAITLTSRQLRTLNVAOELSRGSGTGRVVISAPQGLEVVAYIGALQAGRIAY 64
 DB 52 DLTGTSVAMTYRELIDHAREVATHLRHAGVGGQGVLLHPHGLDYLAAPFGCLAGAVAV 111
 QY 65 PLSPVG---GVTDERSDVSLSDFVALITTSAAVDVYQVHARRPGSPSIIENVL 121
 DB 112 PAYPDMNARGQTPVRLAARCAATHTLRRVRETYAADGTGAVGELDGKMLVYE 171
 QY 122 DLDAPNGYTFKED---EYSTAYLQYTSSTRTPGVVMASHQNVNFE--QLMSGFAD 176
 DB 172 DLTGTSSTAMEDPGATFARSLAFLQYTSSTRTPGVVMASHQNVNFE--QLMSGFAD 231
 QY 177 TDGIPPPNSALVSWLPHYHMGVLVIGICAPILGGYPAVLTPSVFLQRPARRMHLASDF 236
 DB 232 -----SGVSWLPPYHMGVLIGILTPVYGGFPAHLMAMPYTFQRPMLMETLSRTG 283
 QY 237 HAFSAAPNFAFELAAARRTTDDMAGRDGLNLTILSGSERVOATIKRFADRFARN 290
 DB 284 ASTVAPNFGEEQCLRRITPAQARAGDLSRWRLALNGAEIPRPTLDRAEYFA 337

RESULT 13
 OS0857 PRELIMINARY; PRT; 1770 AA.
 AC 050857;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
 DE SAFRAMYCIN MX1 SYNTHETASE B.
 GN SAFB.
 OS Myxococcus xanthus.
 OC Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
 OC Myxococcales; Cystobacteriinae; Myxococcaceae; Myxococcus.
 NCBI_TaxID=34;

RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-DM504-15;
 RX MEDLINE-96032006; PubMed-7551044;
 RA Pospiech A., Gluzel B., Bietenhader J., Schupp T.;
 RT "A new Myxococcus xanthus gene cluster for the biosynthesis of the
 antibiotic saframycin Mx1 encoding a peptide synthetase.",
 RT Microbiology 141:1793-1803(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-DM504-15;
 RX MEDLINE-97000395; PubMed-8936303;
 RA Pospiech A., Bietenhader J., Schupp T.;
 RT "Two multifunctional peptide synthetases and an O-methyltransferase
 are involved in the biosynthesis of the DNA-binding antibiotic and
 antitumour agent saframycin Mx1 from Myxococcus xanthus.",

RT antitumour agent saframycin Mx1 from Myxococcus xanthus.",

RESULT	14		
096554			
ID	096554	PRELIMINARY;	PRT; 8243 AA.
AC	096554;		
DT	01-MAY-1999 (TREMBLrel. 10, Created)		
DR	01-MAY-1999 (TREMBLrel. 10, last sequence update)		
DE	01-OCT-2000 (TREMBLrel. 15, last annotation update)		
GN	TYPE I FATTY ACID SYNTHASE.		
OS	FASTI.		
OC	Cryptosporidium parvum.		
CC	Eukaryota; Alveolata; Alveolomplexa; Coccidia; Elmeritida;		
OX	Cryptosporidiidae; Cryptosporidium.		
NCBI_TaxID=5807;			
SEQUENCE FROM N.A.			
STRAIN=KSU-1;			
RA	Zhu G., Marchevka M.J., Woods K.M., Upton S.J., Keithly J.S.;		
RT	"characterization of a type I FAS gene in the parasitic protozoan		
RL	Cryptosporidium parvum."		
Submitted (Aug-1998) to the EMBL/Genbank/DBJ databases.			
EMBL: AF082993; AAC99407.1; -.			
HSSP: P28304; 100R.			
INTERPRO: IPR000255; -.			
INTERPRO: IPR000794; -.			
INTERPRO: IPR000873; -.			
INTERPRO: IPR001227; -.			
INTERPRO: IPR002085; -.			
PFAM: PF00107; adh_2inc; 3.			
PFAM: PF00109; ketoacyl_synth; 3.			
PFAM: PF00501; AMP-binding; 1.			

Query Match	18.8%;	Score 290;	DB 2;	Length 110;
Best Local Similarity	52.8%;	Pred. No. 3.9e-17;		
Matches 57;	Conservative 13;	Mismatches 36;	Indels 2;	Gaps 1;

DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE HYPOTHETICAL 67.9 KDA PROTEIN RV2930.
GN FADD26 OR RV2930 OR MTCY338.19.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=96295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagers K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,
RA Taylor K., Whitehead S., Barrett B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence."
RL Nature 393:537-544(1998).
CC -I- SIMILARITY: TO ENZYMES WHICH ACT VIA AN ATP-DEPENDENT COVALENT
CC BINDING OF AMP TO THEIR SUBSTRATE. STRONG, TO M.LEPPAE MASC AND
CC M.TUBERCULOSIS RV1521.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: 274697; CAA98985.1; -
CC TUBERCULIST; RV2930; -
CC INTERPRO: IPR000873; -
CC PFMW: PF00501; AMP-binding; 1.
CC DR PFMW: PF00501; AMP-binding; 1.
CC KW Hypothetical protein.
CC SEQUENCE 626 AA; 67935 MW; F53F6AF8C9171548 CRC64;
SQ

Query Match 56.4%; Score 869; DB 1; Length 626;
Best Local Similarity 58.5%; Pred. No. 2,1e-62;
Matches 172; Conservative 40; Mismatches 80; Indels 2; Gaps 1;

QY 1 DYEQDMGVAITLRSQILRFLNAQELSRGSGDRVVISAPOGLEYVAYLGALQAG 60
DB 72 DYGSDPKGFADSLTWSQYYSRACIIAELKLCGLPDRVAVLAPQGLEVLAFGLAQG 131
QY 61 RIAVPLSVQGGVTDERSVLSDSPPAIIITSSAVDVOHVARRPESPPIIENDL 120
DB 132 FIAVPLSPQYGIHDDRSVAVLQDSKPAVILITSSVVGIVTLYAASHDQAPVVEVDL 191
QY 121 IDLDAPNGVTFEDEXPRFAYLQYTSGRTPAGVVMASHQNRVNEEOIOMSGYFADTOSI 180
DB 192 IDLDSPQMPAFSRQHTGATLQYTSGRTPAGVIVSTFNIVANTQSMYGFQDPAKI 251
QY 181 PPNNSALVSWLPFYHDMGLVIGICAPILIGGYPAVLTPSPFYQRPARRMHLMASDFHARS 240
DB 252 --PTGVVSWMLPLHYHDMGLIGICAPILVARRRAMLMSPSFLRRARRMQLLATSGRCS 309
QY 241 AAPNAPFELIARRTTDDMAGRDLCNIIITLISGSEVCAATTKRADRFARPANL 294
DB 310 AAPNAPFELIARRTTDDMAGRDLCNIIITLISGSEVCAATTKRADRFARPANL 363

RESULT 5
ID YD45_MYCTU STANDARD; PRT; 521 AA.
AC Q11015;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)

DE HYPOTHETICAL 53.9 KDA PROTEIN RV1345.
GN FADD33 OR RV1345 OR MTCY02810.09.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=96295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagers K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,
RA Taylor K., Whitehead S., Barrett B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence."
RL Nature 393:537-544(1998).
CC -I- SIMILARITY: TO ENZYMES WHICH ACT VIA AN ATP-DEPENDENT COVALENT
CC BINDING OF AMP TO THEIR SUBSTRATE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: 275555; CAA99978.1; -
CC TUBERCULIST; RV1345; -
CC INTERPRO: IPR000873; -
CC PFMW: PF00501; AMP-binding; 1.
CC DR PROSITE: PS00455; AMP BINDING; 1.
CC KW Hypothetical protein; Transmembrane.
CC TRANSMEM 68
CC TRANSMEM 88
CC TRANSMEM 206
CC TRANSMEM 309
CC TRANSMEM 392
CC TRANSMEM 481
CC TRANSMEM 501
CC SEQUENCE 521 AA; 53957 MW; 2D432512077122DB CRC64;
SQ

Query Match 13.9%; Score 214; DB 1; Length 521;
Best Local Similarity 31.6%; Pred. No. 6,8e-10;
Matches 80; Conservative 27; Mismatches 120; Indels 26; Gaps 9;

QY 47 LEVYVAYLGALDAGRIANVPLSPQGGYDTER-SDSVLSDSSPVAIIITSSAVDVOHVAYA 105
DB 65 VELVAIAIGALAGAAVASTILPPVPGVAGANDQRFWADATLTFELIGIVRTVLSQ-----GSYLA 120
QY 106 RRPGESPSPIIIEVDLDDLDAPN-GYTFEKEDVPRFAYLQYTSGRTPAGVVMASHQNRV 164
DB 121 RLRSDVFAVTTIGDLSTAHTNRSATPVASGEP--AVLQGTAGSGTAPRTAILISGCAVLS 178
QY 165 NFEOLMS--GYFADPDGIPPNNSALVSWLPFYHDMGLVIGICAPILIGGYPAVLTPSPFL 222
DB 179 NLRGLNQRGTAAAD-----VCSWMLPLHYHDMGLAF-VLSAALDAGAPLMLATPTAFT 230
QY 223 QRPARRMHLMASDFHARSAPNFAFELA--ARRTTDDMAGRDLCNIIITLISGSEVCAAT 279
DB 231 ASPFRLMLSWLSDSGATMTAAPNFAVNLIGKYARRVSE-----VDLGAALRVTLNGEPPVC 285
QY 280 ATKRRPDRFARF 292
DB 286 DGLTRFAEAMAF 298

RESULT 6
ID TYCC_BACBR STANDARD; PRT; 6486 AA.
AC TYCC_BACBR
DT TYCC_BACBR

RT gramidin S synthetase 2 gene from *Bacillus brevis*;
 RL J. Biochem. 110:111-119(1991).
 RN [3]
 RP SEQUENCE OF 1-143 FROM N.A.
 RC STRAIN=ATCC 9999;
 RX MEDLINE=9008776; PubMed=2477357;
 RA Kraetzschmar J., Krause M., Marahiel M.A.;
 RT "Gramicidin S biosynthesis operon containing the structural genes
 to grsA and grsB has an open reading frame encoding a protein homologous
 to fatty acid thioesterases";
 RL J. Bacteriol. 171:5422-5429(1989).
 RN [4]
 RP SEQUENCE OF 1-15, AND CHARACTERIZATION.
 RC STRAIN=NAGANO;
 RX MEDLINE=92011463; PubMed=1917901;
 RA Kurotsu T., Hori K., Kanda M., Saito Y.;
 RT "Characterization and location of the L-proline activating fragment
 from the multifunctional gramicidin S synthetase 2.";
 RL J. Biochem. 109:763-769(1991).
 CC -1- FUNCTION: THIS PROTEIN IS A MULTIFUNCTIONAL ENZYME, ABLE TO
 ACTIVATE AND POLYMERIZE THE AMINO ACIDS PRO, VAL, ORN AND LEU.
 CC -1- COFACTOR: CONTAINS 4 COVALENTLY BOUND PHOSPHOPANTHETHEINES.
 CC -1- PATHWAY: NON-RIBOSOMAL BIOSYNTHESIS OF THE CYCLIC PEPTIDE
 ANTIOTIC GRAMICIDIN S (D-PHE-PRO-VAL-ORN-LEU)2.
 CC -1- SUBUNIT: LARGE MULTISUBUNIT COMPLEX OF GRSA AND GRSB.
 CC -1- DOMAIN: CONSISTS OF FOUR MODULES, AND HARBOURS A PUTATIVE
 THIOESTERASE DOMAIN AT ITS C-TERMINAL END. EACH MODULE
 INCORPORATES ONE AMINO ACID INTO THE PEPTIDE PRODUCT AND CAN BE
 FURTHER SUBDIVIDED INTO DOMAINS RESPONSIBLE FOR SUBSTRATE
 ADENYLATION, THIOLATION, CONDENSATION (NOT FOR THE INITIATION
 MODULE), AND EPIMERIZATION (OPTIONAL), AND N METHYLATION
 (OPTIONAL).
 CC -1- SIMILARITY: TO OTHER ENZYMES WHICH ACT VIA AN ATP-DEPENDENT
 COVALENT BINDING OF AMP TO THEIR SUBSTRATE.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X61558; CAA3838.1; -;
 DR EMBL: M29703; AA58719.1; -;
 DR EMBL: X15577; CAA33604.1; -;
 DR EMBL: D00938; BAA00778.1; -;
 DR PIR: S20542; YGASG2.
 DR HSSP: P14687; 1AMU.
 DR INTERPRO: IPR000255; -;
 DR INTERPRO: IPR000873; -;
 DR INTERPRO: IPR001031; -;
 DR INTERPRO: IPR001242; -;
 DR PFAM: PF00501; AMP-binding; 4.
 DR PFAM: PF00668; DUF4; 4.
 DR PFAM: PF00975; Thioesterase; 1.
 DR PFAM: PF00550; PP-binding; 4.
 DR PRINTS: PR00154; AMPBINDING; 4.
 DR PROSITE: PS00012; PHOSPHOPANTHETHEINE; 4.
 DR PROSITE: PS00455; AMP BINDING; 4.
 DR PROSITE: PS0075; ACP DOMAIN; 4.
 DR L1ase; Antibiotic diosynthesis; Multifunctional enzyme;
 KW Repeat; Phosphopantetheine; Hydrolase.
 FT INIT_MET 0
 FT REPEAT 466 1043 DOMAIN 1 (PROLINE-ACTIVATING).
 FT REPEAT 1521 2080 DOMAIN 2 (VALINE-ACTIVATING).
 FT REPEAT 2538 3135 DOMAIN 3 (ORNITHINE-ACTIVATING).
 FT REPEAT 3531 4173 DOMAIN 4 (LEUCINE-ACTIVATING).
 FT DOMAIN 975 1042 ACYL CARRIER (ACP).
 FT DOMAIN 2011 2078 ACYL CARRIER (ACP).
 FT DOMAIN 3057 3124 ACYL CARRIER (ACP).
 FT DOMAIN 4095 4162 ACYL CARRIER (ACP).

FT BINDING 1005 1005 PHOSPHOPANTHETHEINE (BY SIMILARITY).
 FT BINDING 2041 2041 PHOSPHOPANTHETHEINE (BY SIMILARITY).
 FT BINDING 3087 3087 PHOSPHOPANTHETHEINE (BY SIMILARITY).
 FT BINDING 4125 4125 PHOSPHOPANTHETHEINE (BY SIMILARITY).
 FT CONFLICT 274 274 H -> D (IN REF. 2).
 FT CONFLICT 418 418 K -> R (IN REF. 2).
 FT CONFLICT 654 664 NAVILCVKRF -> TCSFDVCYOEI (IN REF. 2).
 FT CONFLICT 941 946 HVRLHL -> OLPLTP (IN REF. 2).
 SQ SEQUENCE 4451 AA; 510036 MW; E0029C9B51F5A4B7 CRC64;
 Query Match 12.2%; Score 187.5; DB 1; Length 4451;
 Best Local Similarity 26.2%; Pred. No. 1.5e-06;
 Matches 79; Conservative 45; Mismatches 116; Indels 61; Gaps 14;
 QY 6 WDGAATITRSQLYRRTLNVAQELSRCSGTDR-VVISAQGLEVVAVYALQAGRIAY 64
 DB 3610 WEGQA-LIYHELNLIKANQALHVRKEGVTTPHVPALMTERLEMIYGFISLKAGAVY 3647
 QY 65 PL--SVPGGVTDESDSVLSPPVAILITSSAVDDVQVHARRPGSPSILEVDLLD 122
 DB 3668 PLDPAYVPQ---ERIQYLLEDSGAALLLTOS-----HYLNK-----LPVDIEW 3706
 QY 123 LDAPNGYTFKED-----EYPTAVLYQYTGSTRPAGVYMSQNVNF-----EQL 169
 DB 3707 LDLTQEQNVVEGCTLPRFNQSTDLAYITTSYGTGKPKGVNIEHST-INCLQKRKEEY 3765
 QY 170 MSGYFADTDGILPPNSALVSWLFFYHDMGLVIGICAPILGGYPAVLTSVPSFLQPARMM 229
 DB 3766 ERG-----PDDTALQVSPFAD-GFVASLFPALAGATSVLPREEE-AKRPVALK 3813
 QY 230 HLMASD--FHAESAAPNFAFEIARRTDDMAGDGLNLTILSGSERVOATTKRPAD 287
 DB 3814 KLIASEELHYGVPSLPSAIL-----DVSSKDLQNLRCVTLGEGKPLQAIYKIKE 3866
 QY 288 R 288
 DB 3867 K 3867
 RESULT 8
 LCFA_ECOLI STANDARD; PRT; 561 AA.
 AC P29212.
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE LONG-CHAIN-FATTY-ACID--COA LIGASE (EC 6.2.1.3) (LONG-CHAIN ACYL-COA
 SYNTHETASE).
 GN FADD OR OLDD.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=94150456; PubMed=8107670;
 RA Fudula M., Heinz E., Wolter F.P.;
 RT "The fadD gene of *Escherichia coli* K12 is located close to rnd at
 39.6 min of the chromosomal map and is a new member of the
 AMP-binding protein family";
 RL Mol. Gen. Genet. 242:241-249(1994).
 RN [2]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-10.
 RC STRAIN=K12;
 RX MEDLINE=93094273; PubMed=1460045;
 RA Black P.N., Dirusso C.C., Metzger A.K., Helmet T.L.;
 RT "Cloning, sequencing, and expression of the fadD gene of *Escherichia*
 coli encoding acyl coenzyme A synthetase.";
 RL J. Biol. Chem. 267:25513-25520(1992).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;

RX MEDLINE-97426617; PubMed-9278503;
 RA Blatner F.R., Plunkett G., III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of *Escherichia coli* K-12.";
 RL Science 277:1453-1474(1997).
 [4]
 RP SEQUENCE FROM N.A.
 RP STRAIN-K12;
 RX MEDLINE-97251358; PubMed-9097040;
 RA Itoh T., Alba H., Baba T., Fujita K., Hayashi K., Inada T.,
 RA Isono K., Kasai H., Kimura S., Kitakawa M., Kitagawa M.,
 RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
 RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T.,
 RA Saito N., Sempel G., Seki Y., Sivasubraman S., Tagami T.,
 RA Takeda J., Takemoto K., Wada C., Yamamoto Y., Horuchi T.;
 RT "A 460-kb DNA sequence of the *Escherichia coli* K-12 genome
 corresponding to the 40,1-50.0 min region on the linkage map.";
 DNA Res. 3:379-392(1996).
 -1- FUNCTION: ESTERIFICATION, CONCOMITANT WITH TRANSPORT, OF EXOGENOUS
 LONG-CHAIN FATTY ACIDS INTO METABOLICALLY ACTIVE COA THIOESTERS
 FOR SUBSEQUENT DEGRADATION OR INCORPORATION INTO PHOSPHOLIPIDS.
 -1- CATALYTIC ACTIVITY: ATP + A LONG-CHAIN CARBOXYLIC ACID + COA = AMP
 + PYROPHOSPHATE + AN ACYL-COA.
 -1- COFACTOR: REQUIRES MAGNESIUM.
 -1- SUBUNIT: HOMODIMER (PROBABLE).
 -1- SUBCELLULAR LOCATION: PARTIALLY MEMBRANE-ASSOCIATED (POTENTIAL).
 -1- SIMILARITY: TO OTHER ENZYMES WHICH ACT VIA AN ATP-DEPENDENT
 COVALENT BINDING OF AMP TO THEIR SUBSTRATE.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: X70994; CAA50321.1; -
 DR EMBL: L02649; AAA23752.1; -
 DR EMBL: A8000275; AAC74875.1; -
 DR EMBL: D90823; BAA15600.1; -
 DR EMBL: D90824; BAA15609.1; -
 DR EMBL: D90825; BAA15614.1; -
 DR PIR: A45062; A45062.
 DR HSP: P08659; ILCI.
 DR ECOCODE: E011530; FADD.
 DR INTERPRO: IPR000873; -
 DR PFAM: PF00501; AMP-binding; 1.
 DR PROSITE: PS00455; AMP-BINDING; 1.
 KW Ligase; Fatty acid metabolism; Magnesium; Membrane.
 FT LIGASE 34 51
 FT FT
 FT CONFLICT 468 490
 FT CONFLICT 496 496
 FT CONFLICT 555 561
 FT SEQUENCE 561 AA; 62332 MW; 249B0A54B3DBFA5 CRC64;
 SO
 Query Match 11.6%; Score 178.5; DB 1; Length 561;
 Best Local Similarity 30.5%; Pred. No. 5.3e-07;
 Matches 68; Conservative 26; Mismatches 78; Indels 51; Gaps 12;
 QY 36 GDRVYISAPQGLEVVAYGALQAGRIAY--PLSVPGGCVTERSDSVLSDSPAAIL- 91
 DB 74 GDRVYISAPQGLEVVAYGALQAGRIAY--PLSVPGGCVTERSDSVLSDSPAAIL- 127
 QY 92 -----TTSAAVD-DYVOAH-ARRPGES-----PSIIIEVDLLDD 124
 DB 128 VSNFAHLEKLVKTAQVAVHILTRMGQSLSTAKGVVNVVYKIKRLPKVHLDPALISFR 187

QY 125 AP--NGYTF--KEDEP-STAYLOYTSGSTRTPAGVMSHONVRVNEQLMSGYFADTD 178
 DB 188 SALHNSQRMQYKPELDELAFLOYTGTTGAVAKAMILTRHMLANLEQVNAVY----- 242
 QY 179 G--IPPNSALVSWLPFYHDMGLVIGTCAPV-IGGTPAVLTSP 218
 DB 243 GPHLHGKELVVALPLVHLFALTINCLLEFELGQGNLLITNP 285
 RESULT 9
 ID PKSL_BACSU STANDARD; PRT; 557 AA.
 AC P40806;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE PUTATIVE POLYKETIDE BIOSYNTHESIS PROTEIN PKSL.
 GN PKSL.
 OS *Bacillus subtilis*.
 OC Bacteria; Firmicutes; *Bacillus*/Clostridium group;
 OC *Bacillus*/Staphylococcus group; *Bacillus*.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-168 / B81424;
 RX MEDLINE-95219083; PubMed-7704258;
 RA Albertini A.M., Carrao T., Scoffone F., Scotti C., Gallizzi A.;
 RT "Sequence around the 159 degree region of the *Bacillus subtilis*
 RT genome: the *pkx* locus spans 33.6 kb.";
 RL Microbiology 141:299-309(1995).
 CC -1- FUNCTION: POTENTIALLY INVOLVED IN SOME INTERMEDIATE STEPS FOR
 CC THE SYNTHESIS OF A POLYKETIDE MOLECULE WHICH MAY BE INVOLVED IN
 CC SECONDARY METABOLISM.
 CC -1- SIMILARITY: TO OTHER ENZYMES WHICH ACT VIA AN ATP-DEPENDENT
 CC COVALENT BINDING OF AMP TO THEIR SUBSTRATE.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: U11039; AAA85143.1; -
 DR EMBL: Z99112; CAB13589.1; -
 DR EMBL: Z99113; CAB13600.1; -
 DR SUBTLIST: BG10929; PKSL.
 DR INTERPRO: IPR000873; -
 DR PFAM: PF00501; AMP-binding; 1.
 DR PRINTS: PR00154; AMP-BINDING.
 DR PROSITE: PS00455; AMP-BINDING; 1.
 KW Antibiotic biosynthesis; Ligase.
 FT SEQUENCE 557 AA; 62063 MW; E173B32386B540E4 CRC64;
 SO
 Query Match 11.5%; Score 176.5; DB 1; Length 557;
 Best Local Similarity 22.9%; Pred. No. 7.7e-07;
 Matches 67; Conservative 53; Mismatches 147; Indels 25; Gaps 7;
 QY 7 DGVAITLTRSOLYRRLNVAQELSRGSGTGDY-VLSAPQGLEVVAYGALQAGRIAYP 65
 DB 54 DGEVYQSYRRIMDDDLRLAKGLRQSLQGLDQNSQLDPAFMCVITGVVAP 113
 QY 66 LSVPGGCVTERSDSVLSDS---SPVAITLTSSADVDDVVOHARRPGSPSIIIEVDLL 121
 DB 114 LAVPPTVYAESSTQKLDKMTLLDKPAVITDPRGAGHQMIDNAKEGLEGFAFIYEDLL 173
 QY 122 DLDAPNGYFKEDEYSTAYLOYTSGSTRTPAGVMSHONVRVNEQL--MSGYFADTDG 179
 DB 174 SAADWDWQSSPE--DLALLLTSGSTGTPKAVMLNHRNIMSVGIIQMGFTRE--- 228
 QY 180 IPPNSALVSWLPFYHDMGLVIGTCAPVILGGTPAVLTSPVSLQRPARMW---HLMASD 235

```

DB 229 -----DITFMMFHDVHGIGMLHLDVYLGCEINVSFTILMEPLKMLWIDHYRAS- 282
QY 236 FIAFSAAPFAPFELARRTDDMAGRDLCNITLITSGSERVOATIKFEAD 287
DB 283 ---VTWAPNFAGCLVTDFAEEIKDKKWLDSMKMYLMNGEAMVAKYGRILE 331

RESULT 10
Y934_HUMAN
ID Y934_HUMAN STANDARD: PRT; 1352 AA.
AC Q9Y2E4;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE HYPOTHETICAL PROTEIN KIA00934 (FRAGMENT).
GN KIA00934.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
[1]
SEQUENCE FROM N.A.
KC TISSUE-BRAIN;
RX MEDLINE-99246063; PubMed-10231032;
RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Hirose M.,
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XIII.
RT for large proteins in vitro."
RL DNA Res. 6:63-70(1999).
CC -1- SIMILARITY: STRONG, TO KIA0184.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL, AB023151; BAA6778.1; -.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 1352 AA; 148575 MW; 6625B91B3A0A2BC6 CRC64;

Query Match 11.1%; Score 170.5; DB 1; Length 1352;
Best Local Similarity 23.5%; Pred. No. 7.4e-06;
Matches 74; Conservative 54; Mismatches 140; Indels 47; Gaps 11;

QY 9 VAITLRSOLYKRTLVNADELSCG--STGDRYVISAPOGLEVVAYLGLAQGRIVAPL 66
DB 794 INSLTLCVLIHKRAEKIAVLMERGLHODGHAVALYPPGIDIAAFYGLVAG--CVPI 851
QY 67 SV-----POGAVPDERSDVLSDDSPVAILITSSAVDDV--OHVARRGESPPSTIEVD 119
DB 852 YRPPPPQNIATILPTVKAIVESRSACLTMTTOLICKLRBAALAVDRTPLILDD 911
QY 120 LLDLPANGYTEKEDYPSYAVLYQYSGTRPAGVYMSH-----QNVRVNEQLDMSG 172
DB 912 DLPKKRP-AQICKPNBDPLAYLDVSVSTGMLAGVKMSHAATSARCSIKLCELY--- 967
QY 173 YFADTGTGIPPSNALSVMLEFHYDMGLVIGICAPILIGYPAVITVSFSLQRRARM--- 229
DB 968 -----PSREVAICDDPYCGILFVIMCLOCSVSGHOSIIPSELETNPALMLAV 1017
QY 230 --HMASDFHAFSAAPNFELAR-----RTDDDMAGRDLGNTL--ITSGSERVOAAT 282
DB 1018 SQKVDVDTCSYS-----VMEICTKGLSGQTESLAKRGDLSIVRCVYVAEERPAIAL 1072
QY 283 KRPADRFARFNLOER 297
DB 1073 QSFSLKFLKDLGLHPR 1087

```

```

RESULT 11
LUCI_PROPY
ID LUCI_PROPY STANDARD: PRT; 550 AA.
AC P08659;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE LUCIFERIN 4-MONOOXYGENASE (EC 1.13.12.7) (LUCIFERASE).
OS Photinus pyralis (North American firefly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga;
OC Elateriformia; Cantharoidea; Lampyridae; Photinus.
[1]
SEQUENCE FROM N.A.
RP MEDLINE-87144243; PubMed-3821727;
RX de Wet J.R., Wood K.V., DeLuca M., Hellinski D.R., Subramani S.;
RT "Firefly luciferase gene: structure and expression in mammalian
RT cells."
RL Mol. Cell. Biol. 7:725-737(1987).
RN [2]
RN SUBCELLULAR LOCATION.
RX MEDLINE-87204117; PubMed-3554235;
RA Keller G.-A., Gould S., de Luca M., Subramani S.;
RT "Firefly luciferase is targeted to peroxisomes in mammalian cells."
RL Proc. Natl. Acad. Sci. U.S.A. 84:3264-3268(1987).
RN [3]
RN X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
RX MEDLINE-96398615; PubMed-8805533;
RA Conti E., Franks N.P., Brick P.;
RT "Crystal structure of firefly luciferase throws light on a
RT superfamily of adenylate-forming enzymes."
RL Structure 4:287-298(1996).
[4]
RN X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).
RX MEDLINE-99007339; PubMed-9788915;
RA Franks N.P., Jenkins A., Conti E., Lieb W.R., Brick P.;
RT "Structural basis for the inhibition of firefly luciferase by a
RT general anesthetic."
RL Biophys. J. 75:2205-2211(1998).
CC -1- FUNCTION: PRODUCES GREEN LIGHT WITH A WAVELENGTH OF 562 NM.
CC -1- CATALYTIC ACTIVITY: LUCIFERIN + O(2) + ATP = OXIDIZED LUCIFERIN +
CC CO(2) + H(2)O + AMP + PYROPHOSPHATE + LIGHT.
CC -1- CORCTOR: REQUIRES MAGNESIUM.
CC -1- SUBCELLULAR LOCATION: PEROXISOMAL.
CC -1- SIMILARITY: TO OTHER ENZYMES WHICH ACT VIA AN ATP-DEPENDENT
CC COVALENT BINDING OF AMP TO THEIR SUBSTRATE.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL, M15077; AAA29795.1; -.
DR EMBL, X84848; CAA59283.1; -.
DR EMBL, U03687; AAA03561.1; -.
DR EMBL, U89934; AAB64396.1; -.
DR EMBL, U89935; AAB64399.1; -.
DR PIR, A26772; A26772.
DR PDB, 1UCI; 26-MAR-97.
DR PDB, 1BA3; 11-NOV-98.
DR INTERPRO: IPR000873; -.
DR PFAM: PF00501; AMP-binding; 1.
DR PROSITE: PS00342; MICROBODIES_CTER; 1.
DR OXIDOREDUCTASE: Monooxygenase; Photoprotein; Luminescence; Magnesium;
KW Peroxisome; 3D-structure.
FT SITE 548 550 MICROBODY TARGETING SIGNAL.
SQ SEQUENCE 550 AA; 60745 MW; E380FC9D56ACCDE CRC64;

```

Query Match	11.0%;	Score 170;	DB 1;	Length 550;
Best Local Similarity	23.5%;	Pred. No. 2.5e-06;		
Matches	74;	Conservative	54;	Mismatches 119;
			Indels	68;
			Gaps	17;

[illegible]

RESULT	12
LCFA_BACSU	
ID	LCFA_BACSU
STANDARD;	PRT;
	560 AA.

DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE LONG-CHAIN-FATTY-ACID-COA LIGASE (EC 6.2.1.3) (LONG-CHAIN ACYL-COA
DE SYNTHETASE).

CC Bacteria; Clostridia
CC Bacteria; Firmicutes; Bacillus/Clostridium group
CC Bacillus/Staphylococcus group; Bacillus.
RN (1)
RP SEQUENCE FROM N.A.

RC SIRAIN=108;
RX MEDLINE=97124191; PubMed=8969504;

Wipacaw, Carter N., Bignell C.S., Guy C.B., Fifer A., Sanders J., Emmerson P.T., Harwood C.R.;

"The dnaB-phea (256 degrees-240 degrees) region of the *Bacillus subtilis* chromosome containing genes responsible for stress responses, the utilization of plant cell walls and primary

Microbiology 142:3067-3078(1996).

+ PYROPHOSPHATE + AN ACYL-COA.

CC - SIMILARITY: TO OTHER ENZYMES WHICH ACT VIA AN ATP-DEPENDENT

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

between the Swiss Institute of Bioinformatics and the EMBL outstation

use by non-profit institutions as long as its content is in

entities requires a license agreement (See <http://www.isb-sib.ch/annou>

DR	EMBL:	Z99118;	CAB14816.1;	-
CC	NCBI:	000000	1000	

DR SUBTILIST; BG1946; LCFA.
NUMBER: TMD000072.

DR PF00501: AMP-binding; 1.
 DR PROSITE; PS00455; AMP_BINDING; 1.
 KW Ligase; Fatty acid metabolism; Magnesium.
 SQ SEQUENCE 560 AA; 62691 MW; 416A664E55226C7EA CRC64

Query Match	10.9%	Score 168;	DB 1;	Length 560;
Best Local Similarity	25.2%	Pred. No. 3.7e+06;		
Matches 82; Conservative	42;	Mismatches 135;	Indels 66;	Gaps 12

QY 7 DGVAIT-----LTRSQLRRTLNVAQELSCGC-STGDRVISAIPQGLEVVAVYLGALQAG 600

Db 39 DKTAISFYGRKLTFFHDIITDALKAFAFLQNGLQKGDGRVAVMLPNCQPRVVISYCVLFAG 980

```

QY      61 RIAV---PLSPVQGGVTDERSDSVLSSSPVALIT-----TSSAVDDVVQVHA 105
      | | | | | : : | | | | | : |
Db     99 GIVVQINPL-----YTEHELEYQLRAQVSVITLDDLEPKAIKMKTYLSIVDQILTSV 152

```

```

0Y 106 RRGESPSTI-----EVDLDD-APNGYTF-----KEDEYPTAYLQ 143
    : | : : | | | | |
Db 153 KDYLFPKNLLYPTQKQVHIDFDKTAANIHTFASCMQKQKTELLTIPKTIPEHDIAYLQ 212

```

Qy	144	YTSSTRFAGVVMSHQNVRNFEQLMSGYFADTDGIPBNSALVSMLPFXYHDMGLVIGI	203
	:	:: :	::: :
Db	213	YTGGTTGAKGVMLTHQNIANTEMCAAMVDYKEGA-----EKVLGIVPFFHYGLTAVM	268

QY 204 CAPILGGYPAVLTSVPSFLQRPARMHIMASDEH--AESAPNFAFELARRTTDDMA 260

Db 269 NYSKLGFEWILLPKFDPLET-----IKIDKHKPTLPFGAPRTIYGLLHH---PELQ 318

QY	261	GRDLGNILITLSSERVQATIKRF	285
		: : :	:
Db	319	HYDLSIKSLSGSALPVEYKQKF	343

RESULT 13
BACA BACIT

ID	BACA_BAC11	STANDARD;	PRI;	5255 AA.
AC	068006;			
DT	15-JUL-1999 (Rel. 38, Created)			
DT	15-JUL-1999 (Rel. 38, last sequence update)			

DE ADENYLYLASE (ILEA) (ISOLEUCINE ACTIVASE); ATP-DEPENDENT CYSTEINE
DE ADENYLYLASE (CYSA) (CYSTEINE ACTIVASE); ATP-DEPENDENT ILEICINE ADENYLYLASE
DE BACITRACIN SYNTHETASE 1 (BA1) [INCLUDES: ATP-DEPENDENT ISOLEUCINE
DE 15-JUL-1999 (REL_38, last annotation update)

DE (LEUO) (LEUCINE ACTIVASE); ATP-DEPENDENT GLUTAMATE ADENYLASE (GLOA)
DE (LEUO) (LEUCINE ACTIVASE); ATP-DEPENDENT GLUTAMATE ADENYLASE (GLOA)
DE (GLUTAMATE ACTIVASE); ATP-DEPENDENT ISOLEUCINE ADENYLASE (ILEA)
DE (ISOLEUCINE ACTIVASE); GLUTAMATE RACEMASE (EC 5.1.1.3)].

05 Bacillus licheniformis.
0C Bacteria: Firmicutes: Bacillus/Clostridium group;
0C Bacillus/Staphylococcus group; Bacillus.
(1)

RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 10716;
RX MEDLINE=98089193; PubMed=9427658;
Went D, Kline A, Schoorndorff R, Marshall M A.

RT "The bacitracin biosynthesis operon of *Bacillus licheniformis* ATCC
10716: molecular characterization of three multi-modular peptide
synthetases,";
RT
doi:10.1007/978-1-4020-0371-0007

CC -1- FUNCTION: ACTIVATES FIVE AMINO ACIDS, INCORPORATES TWO D-AMINO
CC ACIDS, RELEASES AND CYCLIZES THE MATURE BACITRACIN.
CC -1- COFACTOR: CONTAINS 5 COVALENTLY BOUND PHOSPHOPANTETHEINES
CC

CC -1- PATHWAY: NON-RIBOSOMAL BIOSYNTHESIS OF THE CYCLIC PEPTIDE
CC ANTI-BIOTIC BACITRACIN.
CC
CC -1- SUBUNIT: LARGE MULTIEZYME COMPLEX OF BA1, BA2 AND BA3.

CC THE FOURTH MODULE. EACH MODULE INCORPORATES ONE AMINO ACID INTO
CC THE PEPTIDE PRODUCT AND CAN BE FURTHER SUBDIVIDED INTO DOMAINS
CC RESPONSIBLE FOR SUBSTRATE ADENYLATION, THIOLATION, CONDENSATION


```
CC N METHYLATION (OPTIONAL).
CC -I- MISCELLANEOUS: BACITRACIN IS A MIXTURE OF AT LEAST TEN CYCLIC
CC DODECAPEPTIDES, THAT DIFFER BY ONE OR TWO AMINO ACIDS. THE MOST
CC ABUNDANT IS BACITRACIN A, A BRANCHED CYCLIC DODECAPEPTIDE. IT
CC CONTAINS AN N-TERMINAL LINEAR PENTAPEPTIDE MOIETY (ILE-CYS-LEU-D-
CC GLU-ILE) WITH AN ISOLEUCINE-CYSTEINE THIAZOLINE CONDENSATION
CC PRODUCT AND A C-TERMINAL HEPTAPEPTIDE RING (LYS-D-ORN-ILE-D-PHE-
CC HIS-D-ASP-ASN), IN WHICH THE FREE ALPHA-CARBOXY GROUP OF THE C-
CC TERMINAL ASN IS BOUND TO THE EPSILON-AMINO GROUP OF LYSINE. IT
CC CONTAINS FOUR AMINO ACIDS IN THE D-CONFIGURATION (GLU-4, ORN-7,
CC PHE-9, AND ASP-11).
CC -I- SIMILARITY: TO OTHER ENZYMES WHICH ACT VIA AN ATP-DEPENDENT
CC COVALENT BINDING OF AMP TO THEIR SUBSTRATE.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@isb-sdb.ch).
-----
DR EMBL: AF007865; AAC06346.1; -.
DR INTERPRO: IPR000255; -.
DR INTERPRO: IPR000873; -.
DR INTERPRO: IPR001242; -.
DR PFAM: PF00501; AMP-binding; 5.
DR PFAM: PF00668; DUF4; 5.
DR PFAM: PF00550; pp-binding; 5.
DR PRINTS: PR00154; AMP-BINDING.
DR PROSITE: PS00012; PHOSPHOPANTHETHEINE; 3.
DR PROSITE: PS00455; AMP-BINDING; 5.
DR PROSITE: PS50075; ACP_DOMAIN; 5.
DR Ligase: Isomerase: Antibiotic biosynthesis; Phosphopantetheine;
KW Multifunctional enzyme; Repeat.
FT REPEAT 39 612
FT REPEAT 1109 1648
FT REPEAT 2124 2689
FT REPEAT 3164 3732
FT REPEAT 4668 5249
FT REPEAT 621 1037
FT DOMAIN 544 612
FT DOMAIN 1585 1648
FT DOMAIN 2621 2689
FT DOMAIN 3664 3732
FT DOMAIN 5171 5249
FT BINDING 574 574
FT BINDING 1615 1615
FT BINDING 2651 2651
FT BINDING 3684 3684
FT BINDING 5201 5201
FT BINDING 5201 5201
SQ SEQUENCE 5255 AA; 596254 MW; 906EBD68450F85B CRC64;

Query Match 10.5%; Score 162; DB 1; Length 5255;
Best Local Similarity 24.8%; Pred. No. 0.00021;
Matches 77; Conservative 45; Mismatches 124; Indels 64; Gaps 13;

Db 7 DGVATITRSQLYRRLN-----VAQELSRG-STGDRVISAPOGLEVVAYLALQAG 60
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
52 DHTALVFGAQRMTYRELKANTARLREKIGRGSIATIAIDRFEMIIIGITLKAG 111
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
61 RIAVPLSVAGQVTDERSQSVLSDSSPAVLLTSSAVD-----DVGQVHARRPGSSPPS 114
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
112 GAYLPLD-BE--TPKDRIFMLSDTKRAVLLTQKKAADGIDCADIVQ----- 156
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
115 IIEVDLIDAPNGYTFKE-----DEYPSRAYLQYSGSRPAGVVMASHQW-RVNFPG 168
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
157 -----LDKASDGFS-KEPLSSVNDSCDTATITTSSTGIGTGVITPHTSYIRV----- 205
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
169 LMSGYEADTDGIPPNASALVSWLPHYHMGVLGICAPLIGGPAVLTSVPSFLQPARM 228
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
206 -----VQNTNYIDITEDNVILQLSNYSFGDSVDFIGALLNGASLVMIKEKALLN----- 255
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
```

```
QY 229 MHLMASDFAFSAFNF-----AFELARKTTDDMAGRDGLNLTLLSGSERVQAATIKR 284
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 256 INLGSAINEEKRYSVMFTTALFNMTA-----DIHVCLSNLRLKILFGGERASIPHYVKR 309
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 285 FADRFARFNL 294
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 310 VLNHVGRDKL 319
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

RESULT 14
ACSA_BACSU STANDARD: PRT; 572 AA.
ID ACSA_BACSU
AC P39062;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE ACTEYL-COENZYME A SYNTHETASE (EC 6.2.1.1) (ACETATE--COA LIGASE) (ACYL-
DE ACTIVATING ENZYME) (ACETYL-COA SYNTHASE).
GN ACSA.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=95020526; PubMed=7934817;
RA Grundy F.J., Waters D.A., Takova T.Y., Henkin T.M.;
RT "Identification of genes involved in utilization of acetate and
RT acetoin in Bacillus subtilis."
RL MOL. Microbiol. 10:259-271(1993).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=98048467; PubMed=9387221;
RA Lapidus A., Galleron N., Sorokin A., Ehrlich S.D.;
RT "Sequencing and functional annotation of the Bacillus subtilis genes
RT in the 200 kb rmb-dnaB region."
RL Microbiology 143:3431-3441(1997).
CC -I- FUNCTION: ROLE IN GROWTH AND SPORULATION ON ACETATE.
CC -I- CATALYTIC ACTIVITY: ATP + ACETATE + COA = AMP + PYROPHOSPHATE +
CC ACETYL-COA.
CC -I- SIMILARITY: TO OTHER ENZYMES WHICH ACT VIA AN ATP-DEPENDENT
CC COVALENT BINDING OF AMP TO THEIR SUBSTRATE.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@isb-sdb.ch).
-----
DR EMBL: I17309; AAA68287.1; -.
DR EMBL: AF008220; AAC00302.1; -.
DR EMBL: 299119; CAB14946.1; -.
DR PIR: S39646; S39646.
DR HSP: P08659; ILCT.
DR SUBTILIST; BG10370; ACSA.
DR INTERPRO: IPR000873; -.
DR PFAM: PF00501; AMP-binding; 1.
DR PROSITE: PS00455; AMP-BINDING; 1.
KW Ligase: Sporulation.
SQ SEQUENCE 572 AA; 64892 MW; 2E16BB907A253F97 CRC64;

Query Match 10.5%; Score 161.5; DB 1; Length 572;
Best Local Similarity 26.8%; Pred. No. 1.3e-05;
Matches 80; Conservative 41; Mismatches 100; Indels 77; Gaps 17;

QY 27 OEISRCGST-----GDRVISAPOGLEVVAYLALQAGRIAVPL--SVPOGQVTD 75
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 81 EESNRAGNVLRRYGNKKDRVFIEFMRSPDELYFIMLGAIKIGALAGPLFEAFMEGAVKD 140
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
```

Oy	76	ERSDVSLESSPPVAILITSSAVDD-----VOHVARRC--ESPSTIEVD-----	115
Dd	141	R-----LENSKAKVYVTTPRELLERTLPDKLPHLOHVEVVGSEALSGNIIINIDYKAQES	195
Oy	120	-LLDLDAPNGYTEFEDEYPSTAYIQYSGSTRPAGVYMSHOXRVNEFQLMSG-----	172
Dd	196	TRLDIE-----MMDKKDF---LLHYTSGISGTGPKGVLIHHEAM---IOOYOTGKWLDL	244
Oy	173	-----YEADTGDIIPPNNALWSMLPFHDMDLVIGCAPILGCVPAVLTSVPSPFLQPAR	227
Dd	245	KEDDIYWTAD-----PGW-----TGYVGIAPPLMNGATNYI---VGRSPSPES	287
Oy	228	WMHLMAS-DEFAFSAPNFAPFELAAARTTDODMAGR-DLGNIILTISGSERVQAATIK	283
Dd	288	WYGTEIQGVVMWYSAPT-AFRML-MGAGEDMAKAYDLNLSRHVLSVGPBLNDPEVIR	342
RESULT	15		
ACVS_CEPAC	ACVS_CEPAC	STANDARD;	PRT; 3712 AA.
	P25464;		
Dd	01-MAY-1992	(Rel. 22, Created)	
Dt	01-MAY-1992	(Rel. 22, Last sequence update)	
Dt	15-JUL-1998	(Rel. 36, Last annotation update)	
Dd	DELTA-(L-ALPHA-AMINOADIPYL-L-CYSTEINYL-D-VALINE SYNTHETASE		
Dd	(EC 6.-.-.-) (ACV SYNTHETASE) (ACVS).		
GN	PCBAB.		
OS	Cephalosporium acremonium (Acremonium chrysogenum).		
OC	Eukaryota; Fungi; Ascomycota; anamorphic Ascomycota; Acremonium.		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE-91177827; Pubmed-1706706;		
RA	Gutierrez S., Diaz B., Montenegro E., Martin J.F.;		
RT	"Characterization of the Cephalosporium acremonium pcbab gene		
RT	encoding alpha-aminoadipyl-L-cysteinyl-valine synthetase, a large		
RT	multidomain peptide synthetase: linkage to the pcbC gene as a cluster		
RT	of early cephalosporin biosynthetic genes and evidence of multiple		
RT	functional domains.;"		
RL	J. Bacteriol. 173:2354-2365(1991).		
RN	[2]		
RP	PARTIAL SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.		
RC	STRAIN-ATCC 11550;		
RX	MEDLINE-91168300; Pubmed-2076552;		
RA	Hoskins J.A., O'Callaghan N., Queener S.W., Cantwell C.A., Wood J.S....		
RT	Chen V.J., Skatrud P.L.;		
RT	"Gene disruption of the pcbAB gene encoding ACV synthetase in		
RT	Cephalosporium acremonium.;"		
RT	Curr. Genet. 18:523-530(1990).		
	-I- FUNCTION: EACH OF THE CONSTITUENT AMINO ACIDS OF THE TRIPEPTIDE		
	ACV ARE ACTIVATED AS AMINOACYL-ADENYLATES WITH PEPTIDE BONDS		
	FORMED THROUGH THE PARTICIPATION OF AMINO ACID THIOLESTER		
	INTERMEDIATES.		
CC	-I- COFACTOR: CONTAINS 3 COVALENTLY BOUND PHOSPHOPANTETHEINES		
CC	(POTENTIAL).		
CC	-I- PATHWAY: FIRST STEP IN THE BIOSYNTHESIS OF PENICILLIN AND		
CC	CEPHALOSPORIN.		
CC	-I- SIMILARITY: TO OTHER ENZYMES WHICH ACT VIA AN ATP-DEPENDENT		
CC	COVALENT BINDING OF AMP TO THEIR SUBSTRATE.		
CC	PIR: A3851; YGCEVC.		
Dd	HSSDP: P14687; IAMU.		
Dd	INTERPRO: IPRO00255; -		
Dd	INTERPRO: IPRO00873; -		
Dd	INTERPRO: IPRO01031; -		
Dd	INTERPRO: IPRO01242; -		
Dd	PFAM: PF00501; AMP-binding; 3.		
Dd	PFAM: PF00668; DUF4; 3.		
Dd	PFAM: PF00975; Thioesterase; 1.		
Dd	PFAM: PF00550; pp-binding; 3.		
Dd	PRINTS: PR00154; AMPBINDING.		
Dd	PROSITE: PS00012; PHOSPHOPANTETHINE; 2.		
Dd	PROSITE: PS00455; AMP-BINDING; 3.		
Dd	PROSITE: PS00705; ACP_DOMAIN; 3.		
Dd	Llase; Multibiotic biosynthesis; Multifunctional enzyme;		
KW			

KW	Repeat: Phosphopantetheine.	
FT	REPEAT	234 1062
FT	REPEAT	1335 2162
FT	REPEAT	2409 3387
FT	DOMAIN	795 864
FT	DOMAIN	1880 1953
FT	DOMAIN	2960 3027
FT	BINDING	827 927
FT	BINDING	1916 1916
FT	BINDING	2990 2990
FT	ACT. SITE	3568 3568
SO	SEQUENCE	3712 AA; 414767 MW; 4EEC1B5EEB9B7 CRC64;

[illegible]

Search completed: January 5, 2001, 14:31:21
Job time: 352 sec

DR PROSITE: PS00455; AMP_BINDING; 3.
 DR PROSITE: PS50075; ACP_DOMAIN; 3.
 KW Ligase: Antibiotic biosynthesis; Multifunctional enzyme

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 5, 2001, 14:28:34 ; Search time 67.42 seconds
(without alignments)
302.139 Million cell updates/sec

Title: US-09-461-774-6

Perfect score: 1540

Sequence: 1 DYEQDMGVAITLTRSOLYR.....TIKRFADRFARFNLQERVKA 300

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 195891 seqs, 6790655 residues

tal number of hits satisfying chosen parameters: 195891

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

PIR66:*
1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1519	98.6	580	2	B70668	probable Acyl-CoA
2	954.5	62.0	584	2	B70820	probable polyketid
3	948	61.6	584	2	F70522	probable polyketid
4	932	60.5	372	2	S72711	masc protein - Myc
5	929	60.3	583	2	A70723	probable acyl-coas
6	906.5	58.9	578	2	A70877	probable acyl-coas
7	869	56.4	626	2	B70749	probable Acyl-CoA
8	833	53.4	619	2	C70669	probable acyl-CoA
9	604.5	39.3	585	2	C70634	probable fadD30 pr
10	484	31.4	637	2	E70887	probable fadD32 pr
11	449	29.2	4342	2	H83343	probable non-ribos
12	426	27.7	620	2	A70635	probable fadD31 pr
13	421	27.3	2297	2	T34918	probable polyketid
14	386.5	25.1	1770	2	T18551	saframycin Mx1 syn
15	310	20.1	8243	2	T31307	type I fatty acid
16	257.5	16.7	562	2	A70702	probable acyl-CoA
17	232	15.1	544	2	H70603	probable polyketid
18	223	14.5	3971	2	T44806	mycosubtilin synth
19	214	13.9	521	2	H70739	probable polyketid
20	193.5	12.6	4450	2	JX0340	gramicidin S synth
21	193.5	12.6	7463	2	T36248	CDA peptide synthet
22	190.5	12.4	6486	2	T31076	tyrocidine synthet
23	190	12.3	4976	2	T14165	peptide synthetase
24	187.5	12.2	4452	1	TG8SG2	gramicidin S synth
25	185	12.0	3643	2	T36410	probable polyketid
26	182.5	11.9	525	2	H70982	probable fadD7 pro
27	181.5	11.8	584	2	C75364	lysobactin synthet
28	181.5	11.8	1575	2	T18545	long-chain-fatty-a
29	176.5	11.6	561	1	S41589	

30	177.5	11.5	8563	2	T30226	polyketide synthas
31	177	11.5	3670	2	T36249	CDA peptide synthet
32	176.5	11.5	557	2	H69678	involved in polyke
33	176.5	11.5	9376	2	T14593	stryngomycin synth
34	173	11.2	7576	2	T17428	FK506 polyketide s
35	172.5	11.2	2591	2	T30288	pristinamycin I sy
36	172	11.2	1394	2	T34061	hypothetical prote
37	171.5	11.1	2611	2	T14591	actinomycin synthet
38	170.5	11.1	552	2	E69438	probable fatty-act
39	170.5	11.1	879	2	B70014	antibiotic synthet
40	170.5	11.1	4077	2	T17484	hypothetical prote
41	170	11.0	550	1	A26772	Photinus-luciferin
42	170	11.0	1809	2	T17403	pyochelin synthet
43	170	11.0	1809	2	C83118	pyochelin synthet
44	168.5	10.9	547	2	A70551	probable acid-CoA
45	168	10.9	560	2	D69649	probable long-chain

ALIGNMENTS

RESULT 1
B70668
probable Acyl-CoA Synthetase - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence-revision 17-Jul-1998 #text-change 18-Aug-2000
C:Accession: B70668
R:COLE, S.T.; BROSCHE, R.; PARKHILL, J.; GARNIER, T.; CHURCHER, C.; HARRIS, D.; GORDON ; CONNOR, R.; DAVIES, R.; DEVLIN, K.; FELTWEILL, T.; GENIES, S.; HUMLIN, N.; HOLROYD, Rajandream, M.A.; ROGERS, J.; RUTER, S.; SEEGER, K.; SKELTON, S.; SQUARES, S. Nature 393, 537-544, 1998
A:Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno
A:Reference number: A70500; MUID:98295987
A:Accession: B70668
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-580 <COL>
A:Cross-references: GB:283858; GB:AL123456; NID:g3261675; PIDN:CAR06100.1; PID:g17811
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: fadD28
C:Superfamily: 4-coumarate--CoA ligase; acetate--CoA ligase homology
F:58-566/Domain: acetate--CoA ligase homology <ACL>

Query Match 98.6%; Score 1519; DB 2; Length 580;
Best local similarity 99.3%; Pred. No. 1.3e-115;
Matches 296; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY	1	DYEQDMGVAITLTRSOLYRRLNVAOELSRCSTGDRVISAPOGLEVVAVIALGALQAG	60
DB	27	DYEQDMGVAITLTRSOLYRRLNVAOELSRCSTGDRVISAPOGLEVVAVIALGALQAG	86
QY	61	RIAVPLSVQGGVTDERSDSVLSDDSPVALITLTSSAVDDVQVAVARRGESPSSIIEVLD	120
DB	87	RIAVPLSVQGGVTDERSDSVLSDDSPVALITLTSSAVDDVQVAVARRGESPSSIIEVLD	146
QY	121	LDLDAANGYTFKDEYPSATVAYLOYTSGSTRTPAGVMSHONVNFEOQLMSGCFADTDGI	180
DB	147	LDLDAANGYTFKDEYPSATVAYLOYTSGSTRTPAGVMSHONVNFEOQLMSGCFADTDGI	206
QY	181	PPNSALVSMLEPYHDMGLVIGICADILGYPVALTSPVSFLQRPARMHMLMSDFHAES	240
DB	207	PPNSALVSMLEPYHDMGLVIGICADILGYPVALTSPVSFLQRPARMHMLMSDFHAES	266
QY	241	AAANFAPFLAARTTDDDMAGRDLGNILTLSSSERVQAAATIKRDFRAREFNLQERV	298
DB	267	AAANFAPFLAARTTDDDMAGRDLGNILTLSSSERVQAAATIKRDFRAREFNLQERV	324

RESULT 2
B70820

[illegible]

RESULT 11
 HB3343
 Probable non-ribosomal peptide synthetase PA2424 [Imported] - *Pseudomonas aeruginosa* (Strain PA01)
 C.Species: *Pseudomonas aeruginosa*
 C.Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 15-Sep-2000
 C.Accession: HB3343
 S.Stover, C.K.; Plam, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Barradine, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lardig, K.; Lim, J.; Lorry, S.; Olson, M.V.
 Nature 406, 959-964, 2000
 A.Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen
 A.Reference number: AB2950
 A.Accession: HB3343
 A.Status: Preliminary
 A.Molecule type: DNA
 A.Residues: 1-4342 <STO>
 A.Cross-references: GB:AE004669; GB:AE004091; NID:g9948460; PIDN:AA05812.1; GSPDB:GN000101
 A.Experimental source: Strain PA01
 C.Genetics:
 A.Gene: PA2424

Query Match	Similarity	29.2%	Score	449	DB	2	Length	4342
Best Local	Similarity	35.4%	Pred	No	1.5e-27			
Matches	103	Conservative	49	Mismatches	129	Indels	10	Gaps
QY	7	DGVAITLRSQLYRRTLVAOELSRGSGTGRRVVISAPQGLEVVAYVIGALOAGRIAVPL	66					
Db	35	DOEGVSLSTRDIDLBARSIQAALQAHQAQGLGRDAVLLFFPSGDPYVAFAFGCLYAGVIAVPA	94					
QY	67	SVPOGVT--DERSBSVLSDSPVALILFTSSAVDDVVOHVARPGESPSTIEVYLLDD	124					
Db	95	YPPESARRHHQRRLLSIADAPERVLVTTADLREFLLQNMQLSANNPQLLCVQDLDA	154					
QY	125	APNGTTFKEDLPSTAYVLYQYTSGRTRTPAGVYMSQONRVNFEQJMSGYFADTDGIPPPN	184					
Db	155	VAEADDEPVRDEHIAFLQYTSGLALPGQVSHSGNLVANNVLLRRFGIGADVV----	210					
QY	185	SALVSMLPFPYHMGIVIGICAPILGYPAYLTSPPVSFLQRPARMHMLASDFHAFSAAPN	244					
Db	211	--IVSMLPFIYHMGILIGLQIPFIQSGVPCVLSMPRTFLERPVRMLIALISQYCGYTSQGPD	268					
QY	245	FAFELAAARTTDDMAGRDLGNILITLLSSERVQAATIKRFADRA--RFN	293					
Db	269	FAYRLCSSEVAASALQRLDLSGMRVAFSGSEPIRQDSLERFKEKPAASFFD	319					

RESULT 12
A70635
probable fadD31 protein - Mycobacterium tuberculosis (strain H37Rv)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 18-Aug-2000
C:Accession: A70635
R:Cole, S.T.; Broesch, R.; Parthill, J.; Garner, T.; Churcher, C.; Harris, D.; Gordon
Comor, R.; Davies, R.; Devlin, K.; Fellwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
Rajadream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulstrom, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:96295987
A:Accession: A70635
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-620 <COL>
A:CROSS-references: GB:264498; GB:AL123456; NID:q3261701; PIDN:CAB06501.1; PID:e293666
A:Experimental source: strain H37Rv
C:Genetics:
A:Gene: fadD31
C:Superfamily: acetate--CoA ligase homology
P:88-608/Domain: acetate--CoA ligase homology <ACL>

Query Match 27.7%, Score 426; DB 2; Length 620;
Best Local Similarity 35.9%, Pred. No. 7.9e-27;
Matches 108; Conservative 55; Mismatches 120; Indels 18; Gaps 8;

```
OY      1 DYEDMDGVAITLTRSOLRYRTLNVQELSRKSGTGRVVISAPDGLFYVAVLGAIALAG 60  
          : : : : : : : : : : : : : : : : : : : : : : : : : : : :  
DB      57 DHARSACCALEVTWTQCGMRLLAAGAIVORFAGCDREVALILAPOGDIDYVCGFYAAIATAG 116  
  
OY      61 RIAYPLSPGGGVNDERSDSVLSDDSPVAIIITTSASVDVVQHAVRRGPESPSIEIDL 126  
          : : : : : : : : : : : : : : : : : : : : : : : : : : : :  
DB     117 TVAAPLFPELPBGHAERLDLTALROSEPAVLIITTAAKAAVEGFNNVRPLRKPTVLVIDQ 176  
  
OY     121 LDLAPANGVFTEKEDEPSTAYLOLYNSGSTRTPAQVAMSHOVNRFNEQLMSGYFAADTDGI 180  
          : : : : : : : : : : : : : : : : : : : : : : : : : : : :  
DB     177 IPDRGEGLFVPEMIDDAVSHLQYTSGSTRPVGVEITHRAVGTNLVQMILSI-----DL 231  
  
OY     181 PPPNALYSWLPFTYHDMGL-VIGICAPILGGYPANLVTPSPFLQRPARMMHLMSDF--- 236  
          : : : : : : : : : : : : : : : : : : : : : : : : : : : :  
DB     232 LNRNHGVSMLPLXHDMLSMIGPPA-VYGGH-STLMSPTLFVRRPLRMIALISGSRTFG 289  
  
OY     237 HAFSAVPNAFEELARR--TTDDMAGRDLGNILTIISGESERVQAATIKRPFADFARN 293  
          : : : : : : : : : : : : : : : : : : : : : : : : : : : :  
DB     290 RVTATAAPNFATVMAAQRLGPAGDDV---DLSNVLLII-GSEPVSIDAVTTFNKAFAFYG 345  
  
OY     294 L 294  
          |  
DB     346 L 346
```

RESULT 13
T34918
polyketide synthase - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C>Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 28-Jul-2000
C:Accession: T34918
R:Oliver, K.; Harris, D.; Parkhill, J.; Barrall, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, January 1998
A:Reference number: Z21558
A:Accession: T34918
A>Status: Preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2297
A:Cross-references: EMBL:AL021409; PIDN:CAA16183.1; GSPDB:GN00070; SCOEDB:SC3F7.12
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCOEDB:SC3F7.12
C:Superfamily: 3-oxoacyl-[acyl-carrier-protein] synthase I homology; [acyl-carrier-pr

GenCore version 4.5
 Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 5, 2001, 14:27:24 ; Search time 45.19 seconds
 (without alignments)
 119.210 Million cell updates/sec

Title: US-09-461-774-6

Perfect score: 1540
 Sequence: 1 DYEDMDGVAITLRSQLYR.....TIKRFADRFARFNQGEVKA 300

Scoring table: BLOSUM62
 Gapop 10.0 , Gapext 0.5

Searched: 174772 seqs, 17957048 residues

tal number of hits satisfying chosen parameters: 174772

Minimum DB seq length: 0
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database : Issued Patents-AA:*
 1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
 2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
 3: /cgn2_6/ptodata/2/1aa/6.COMB.pep:*
 4: /cgn2_6/ptodata/2/1aa/PTUS.COMB.pep:*
 5: /cgn2_6/ptodata/2/1aa/Backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query	Length	ID	Description
1	172	11.2	561	US-08-474-169-8	Sequence 8, Appli
2	171	11.1	550	US-08-718-425-5	Sequence 5, Appli
3	170	11.0	549	US-08-354-240A-2	Sequence 2, Appli
4	170	11.0	550	US-08-867-352-23	Sequence 23, Appli
5	170	11.0	550	US-08-718-425-2	Sequence 2, Appli
6	170	11.0	815	US-08-122-520C-9	Sequence 9, Appli
7	168	10.9	550	US-08-354-240A-4	Sequence 4, Appli
8	168	10.9	550	US-08-354-240A-6	Sequence 4, Appli
9	161.5	10.5	3712	US-08-222-617A-13	Sequence 13, Appli
10	160.5	10.4	3665	US-08-222-617A-13	Sequence 13, Appli
11	160.5	10.4	3712	US-08-222-617A-25	Sequence 25, Appli
12	144	9.4	1410	US-09-335-409-3	Sequence 3, Appli
13	143	9.3	798	US-08-222-617A-8	Sequence 8, Appli
14	139	9.0	15281	US-08-471-119A-2	Sequence 2, Appli
15	129	8.4	739	US-08-510-646B-33	Sequence 33, Appli
16	128.5	8.3	669	US-07-861-800-2	Sequence 2, Appli
17	124	8.1	552	US-09-111-752-5	Sequence 5, Appli
18	124	8.1	797	US-08-222-617A-9	Sequence 9, Appli
19	119	7.7	768	US-08-222-617A-5	Sequence 5, Appli
20	119	7.7	3666	US-08-222-617A-12	Sequence 12, Appli
21	119	7.7	3727	US-08-222-617A-27	Sequence 27, Appli
22	119	7.7	3778	US-08-222-617A-2	Sequence 2, Appli
23	116.5	7.6	552	US-08-231-729B-6	Sequence 6, Appli
24	114	7.4	552	US-09-111-752-7	Sequence 7, Appli
25	111	7.2	548	US-07-963-211-2	Sequence 2, Appli
26	111	7.2	548	US-07-903-047-2	Sequence 2, Appli
27	111	7.2	548	US-08-076-042-P	Sequence 2, Appli
28	111	7.2	548	US-09-111-752-14	Sequence 14, Appli

29	111	7.2	1336	2	US-08-551-356-6	Sequence 6, Appli
30	111	7.2	1336	4	PCT-US93-12687-6	Sequence 6, Appli
31	111	7.2	2324	1	US-08-283-857-1	Sequence 1, Appli
32	111	7.2	2324	1	PCT-US93-09819-1	Sequence 1, Appli
33	111	7.2	2327	5	5455158-1	Patent No. 5455158
34	111	7.2	2386	2	US-09-016-366A-12	Sequence 12, Appli
35	111	7.2	2446	4	US-08-551-356-2	Sequence 2, Appli
36	111	7.2	2446	4	PCT-US93-12687-2	Sequence 2, Appli
37	108	7.0	161	2	US-08-403-852D-30	Sequence 30, Appli
38	108	7.0	161	3	US-08-510-646B-31	Sequence 31, Appli
39	108	7.0	548	2	US-08-460-934-2	Sequence 2, Appli
40	108	7.0	548	2	US-08-782-118-2	Sequence 2, Appli
41	108	7.0	552	3	US-09-111-752-10	Sequence 10, Appli
42	108	7.0	568	2	US-08-460-934-6	Sequence 6, Appli
43	108	7.0	568	2	US-08-782-118-6	Sequence 6, Appli
44	108	7.0	636	2	US-08-460-934-9	Sequence 9, Appli
45	108	7.0	636	2	US-08-782-118-9	Sequence 9, Appli

ALIGNMENTS

```

RESULT 1
US-08-474-169-8
; Sequence 8, Application US/08474169
; Patent No. 5851796
; GENERAL INFORMATION:
; APPLICANT: Schatz, David G.
; TITLE OF INVENTION: An Autoregulatory Retracycline-Regulated
; SYSTEM OF INVENTION: System for Inducible Gene Expression in Eucaryotes
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C
; STREET: 1100 New York Ave. Suite 600
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/474,169
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 561 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-474-169-8

Query Match 11.2%; Score 172; DB 2; Length 561;
Best Local Similarity 24.7%; Pred No 7.5e-10;
Matches 77; Conservative 59; Mismatches 114; Indels 62; Gaps 18;

OY 9 VAITLRSQLYRTLNVAQELSRG-STGDRVYISAPDGLLEVAVYALGAGRIAPVLS 67
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 58 IEVNITAYAYEFEMSVLAELAVKRYGNTNHRITVPEFNSLQFMPVYLGALFV-VAV--- 113
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 68 VQGGVTDRS-DYSVSDSPAAILTTSSAVDDVVOHVARREESPPTIIEVDLDA- 125
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 114 APANDYNERELINSNISQPNVYLSKGLKIL-NVOKK---LPIIOKIIIMDKTD 168
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 126 ---PGYFKEDYV-----STAYLYTGSGSTFTPGVYMSHOWNV 164
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
  
```


Db 215 ACVRFSHARDPIFGNQ---IIPDAII,LSVFPFHGGFGMTTGLYIC-----GPRVYL-- 264

Qy 218 PVSFLOPARAMHMLA--SDFAFS- -APNFAEFLARTRTDDDMAGRDIGNITITLSG- 273

Db 265 ----MYREBELFLRSLDQYKIQSALLVPTL-FSPFAKSTLIDKY---DLSNLEHISGG 316

Qy 274 ---SERVOATITKRP 285

Db 317 APLSKEVGEAAVAKRP 331

```

1      RESULT          4
2      US-08-867-352-23
3      : Sequence 23, Application US/08867352
4      : Patent No. 6080273
5      : GENERAL INFORMATION:
6      : APPLICANT:
7      : TITLE OF INVENTION: Multicistronic expression units and their use
8      : NUMBER OF SEQUENCES: 25
9      : COMPUTER READABLE FORM:
10     : MEDIUM TYPE: Floppy disk
11     : COMPUTER: IBM PC compatible
12     : OPERATING SYSTEM: PC-DOS/MS-DOS
13     : SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPA)
14     : CURRENT APPLICATION DATA:
15     : APPLICATION NUMBER: US/08/867,352
16     : FILING DATE:
17     : CLASSIFICATION:
18     : PRIOR APPLICATION DATA:
19     : APPLICATION NUMBER: 08/387,847
20     : FILING DATE:
21     : INFORMATION FOR SEQ ID NO: 23:
22     : SEQUENCE CHARACTERISTICS:
23     : LENGTH: 550 amino acids
24     : TYPE: amino acid
25     : TOPOLOGY: linear
26     : MOLECULE TYPE: protein
27     :
28     : US-08-867-352-23

```

	Query Match	11.0%;	Score 170;	DB 3;	Length 550;		
	Best Local Similarity	23.5%;	Pred. No. 1.2e-09;				
	Matches	74;	Conservative	54;	Mismatches 119;	Indels	Gaps
Oy	9 VAITLTRSOLYRRTLNVAQELSRCG-STGDRIIVAPOGLEVVAYVAGLQAGRIAVPLS	67					
Dd	47 IEVTNTEAEFMSVRILAEAMKRYGLNTNHRIRLVQSENSLSGFPMVLGLFFIG-VAV---	102					
Dd	68 VPQGCVTIDERS-DSVLSDSSPVAILTTSSAUNDVQNHARRP-----	108					
Dd	103 APANDIYNRELLINSMNISOPFTVNVFSKKLOKILINVOKPIEIOKIIMDSKTDYQQGO	162					
Oy	109 -----GESPSIIIEVDLDLDAANGCYTFKRDELPSTALYDTGGSTTPAGVVMASHN	161					
Dd	163 SMYTEVTSHLRPGENEYDVP-----PEEF---DRDKTIALLINNSSSGSTSLPRGVALLPHPT	214					
Oy	162 VVFNFEEQLMSGFYATPDGIIPRNLSALVSWLRFYHMDGL--VYG--ICAPILGCPVALRLS	217					
Dd	215 ACVRSRHADRPFGNQ-----IIDPTALISVPPRHNGFMGTTLGYLIC-----GFRVYL--	264					
Oy	218 PVSFLQRRPARMWHMLMA-SDFHAESA-AENFAFEILAARRTDDDMAGRDLGNILTLLSG	273					
Dd	265 ---MYREFEELFLRSLDYUKIQSALLVPTY-LFSFPAAKTLLDKV--DLSNLHEIASGG	316					
Oy	274 ----SERVOAQATIKRF	285					
Dd	317 APLSKVEGEAVAAKRK	331					

RESULT 5
US-08-718-425-2
; Sequence 2, Application US/08718425
; Patent No. 6132983

```

1 GENERAL INFORMATION:
2 APPLICANT: Lowe, Christopher R.
3 APPLICANT: White, Peter J.
4 APPLICANT: Murray, James A. H.
5 APPLICANT: Squitrell, David J.
6 TITLE OF INVENTION: LUCIFERASES
7 NUMBER OF SEQUENCES: 5
8 CORRESPONDENCE ADDRESS:
9 ADDRESSEE: Nixon & Vanderhye
10 STREET: 1100 No. 6132983th Glebe Road, 8th floor
11 CITY: Arlington
12 COUNTRY: U.S.A.
13 ZIP: 22201-4714
14
15 COMPUTER READABLE FORM:
16 MEDIUM TYPE: Floppy disk
17 COMPUTER: IBM PC compatible
18 OPERATING SYSTEM: PC-DOS/MS-DOS
19 SOFTWARE: Patentin Release #1.0
20 CURRENT APPLICATION DATA:
21 APPLICATION NUMBER: US/08/718,425
22 FILING DATE: 20-NOV-1996
23 CLASSIFICATION: 435
24 ATTORNEY/AGENT INFORMATION:
25 NAME: Arthur R. Crawford
26 REGISTRATION NUMBER: 25,327
27 REFERENCE/DOCKET NUMBER: 124-539
28 TELECOMMUNICATION INFORMATION:
29 TELEPHONE: (703) 816-4000
30 TELEFAX: (703) 816-4100
31 INFORMATION FOR SEQ ID NO: 2:
32 SEQUENCE CHARACTERISTICS:
33 LENGTH: 550 amino acids
34 TYPE: amino acid
35 STRANDEDNESS: single
36 TOPOLOGY: unknown
37 MOLECULE TYPE: protein
38
39 US-08-718-425-2

```

[illegible]

```

RESULT      6
US-08-122-520C-9
; Sequence 9, Application US/081225200
; Patent No. 5639663
; GENERAL INFORMATION:
;

```



```

ADDRESSER: Demitt Ross & Stevens, S.C.
STREET: 8000 Excelsior Drive, Suite 401
CITY: Madison
STATE: WI
COUNTRY: USA
ZIP: 53717-1914
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/354,240A
FILING DATE: 12-DEC-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Sara, Charles S.
REGISTRATION NUMBER: 30,492
REFERENCE/DOCKET NUMBER: 34506,029
TELECOMMUNICATION INFORMATION:
TELEPHONE: 608-831-2100
TELEFAX: 608-831-2106
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 550 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-354-240A-6

```

```

Query Match          10.9%: Score 168; DB 1: Length 550;
Best Local Similarity 23.5%: Pred. No. 1.9e-09;
Matches 74; Conservative 53; Mismatches 120; Indels 68; Gaps 17;

```

```

QY 9 VAITLRSOLYRRTLNVAQELSRG--STGDRVVISAPQGLEVVAVYALGALGRIAPLVS 67
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 48 IEVDTVAEYFEMSVRLAEAMKRYGINTNHRIVCSSENSLOFMVVLGFLFG-VAV--- 103
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 68 VQGVGTDBRS-DVSLSDSSPAVAILTTSSAVDDVOHVARRP----- 108
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 104 AAANDIYNREELNSMGISQPTVFVSKKGLQKILNVOKKPLIIQKIIIMDSKTDYQGF 163
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 109 -----GESPSITIEVDLDDAPNGYFRKDEYFSTAYLYTGSGSTTPAGVMSHON 161
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 164 SWYTVTSHLPFGENEYDVP---PESE---DRDKTIALINSSSGSTGLPRGVALLPRT 215
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   162 VVAVNEQLMSGYFADTDGIPPNASALVSWLPFYHDMGL--VIG--ICAPILGYPVAVLTS 217
       : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   216 ACVRSHADPFLFGNO---IIPDTAILSVYPRFHGFGMTTILGYLIC-----GFRVYL-- 265
       : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 218 PVSFLQRPARMHMLA-SDFHAFSA--APNFAFELAAARRTTDDMAGRDGLGNITLILSG- 273
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 266 ---MYREFEELFLRSLOQYKIQSALLVPTL-FSFFAKSTLIDKY--DLSMLHEIASG 317
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 274 ---SERVQAATIKRF 285
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 318 APLSKEVGEAVAKRF 332
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 9
US-08-222-617A-4
Sequence 4, Application US/08222617A
Patent No. 5882879
GENERAL INFORMATION:

```

```

APPLICANT: Veenstra, Annemarie E.
APPLICANT: Martin, Juan F.
APPLICANT: Garcia, Bruno D.
APPLICANT: Gutierrez, Santiago
APPLICANT: Barredo, Jose L.
APPLICANT: Von Doehren, Hans
APPLICANT: Palissa, Harriet
APPLICANT: Van Liempt, Henk

```

```

APPLICANT: Montenegro, Eduardo P.
TITLE OF INVENTION: A Method for Influencing Beta-lactam
TITLE OF INVENTION: Antibiotic Production and for Isolation of Large
TITLE OF INVENTION: Quantities of Acv Synthetase
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
STREET: 300 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/222,617A
FILING DATE: 04-APR-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
REFERENCE/DOCKET NUMBER: 97,157
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 3712 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
FEATURE:
NAME/KEY: Modified-site
LOCATION: 2555
OTHER INFORMATION:
OTHER INFORMATION: /note= "Xaa-Ala or Ser"
US-08-222-617A-4

```

```

Query Match          10.5%: Score 161.5; DB 2: Length 3712;
Best Local Similarity 22.8%: Pred. No. 2.1e-07;
Matches 69; Conservative 55; Mismatches 129; Indels 49; Gaps 11;

```

```

QY 7 DGVAI-----TLTRSQLYRRTLNVAQELSRG--TGDRVVISAPQGLEVVAVYALGALOA 59
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 2450 DKAIADAGTGRSLSYSELNRAQVHLIISASIAADDRIALLLKSDIMVALLAWYKA 2509
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 60 GRIAAVPLSVQGVTDERSDVSLSDSSPAVAILTTSSAVDDVOHVARRPGESPSIIEVD 119
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 2510 GAAVYPLDPT---YPSQRELLIESSARTLITTR-----KHNP-----GTVANVP 2554
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 120 LLDLAPN-----GYTFKDEYFST---AYLQYTGSGSTTPAGVMSHONVAVNEEQ 168
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 2555 XVVLDSPEFLACLNOOSKENPTSTQKPSDLAVYVFTSGTGKPGVAVLHOSVYVOLRNS 2614
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 169 LMSGYFADTDGIPPNASALVSWLPFYHDMGLVIGCAPILGYPVAVLTSVPSFLQRPARM 228
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 2615 LTERYFGETNG---SHAVLFISNIVFDSLE-QCLSVLGSNKLIL--PPEEGLTNEAFY 2668
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 229 MHLMASDFHAFSAAPNFAFELAAARRTTDDMAGRDGLGNITLILSGSERVQAATIKRFADR 288
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 2669 DIGREKLSYLSCTPSVLOQIELSR-----LPHLMWYTAGDEFFHNSQFEKIHNSQ 2718
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 289 FA 290
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 2719 FA 2720
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 10
US-08-222-617A-13
Sequence 13, Application US/08222617A
Patent No. 5882879
GENERAL INFORMATION:
APPLICANT: Veenstra, Annemarie E.
APPLICANT: Martin, Juan F.

```

APPLICANT: Garcia, Bruno D.
APPLICANT: Gutierrez, Santiago
APPLICANT: Barredo, Jose L.
APPLICANT: Von Doehren, Hans
APPLICANT: Palissa, Harriet
APPLICANT: Montenegro, Eduardo P.
TITLE OF INVENTION: A Method for Influencing Beta-Lactam
TITLE OF INVENTION: Antibiotic Production and for Isolation of Large
TITLE OF INVENTION: Quantities of ACV Synthetase
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
STREET: 300 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/222,617A
FILING DATE: 04-APR-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
REFERENCE/DOCKET NUMBER: 97,157
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 3665 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Acremonium chrysogenum
FEATURE:
NAME/KEY: Protein
LOCATION: 1..3665
OTHER INFORMATION: /label= ACVS
OTHER INFORMATION: /note="ACV Synthetase from Acremonium
OTHER INFORMATION: chrysogenum; aa 1-3665"
US-08-222-617A-13

Query Match 10.4%; Score 160.5; DB 2; Length 3665;
Best Local Similarity 22.8%; Pred. No. 2.6e-07;
Matches 69; Conservative 55; Mismatches 129; Indels 49; Gaps 11;
DB 7 DGVAI-----TLRSQLYRRTLVNAQELSRGCS--TGDRVISAPOGLEVVAYVYALGALQA 59
2450 DRIADGTRSLSYSELNERANQLVHLIISASIVADDRIALLDKSIDVIALVAMKA 2509
DB 60 GRIAVPLSPQGVTDERSDVSLSPPVAILTSSAVDDVOHVHARRCESPSTIEVD 119
2510 GAAYVPLDPT--YPSQRELIIEESSARTLTTR-----KHPR-----GQIVANVP 2554
DB 120 LLDLDAPN-----GYTFKDEYVST-----AYLYTSGSTRPAGVVMHONVRNFEQ 168
2555 SVYLDSPETLACINQOSKEPPTSTOKPSDLAVITFSGTGKPKGVLVHOSVQLRNS 2614
DB 169 LMSGYADTDGIPPPNSALVSWLPFYHDMGLVIGICAPILIGYPAVLTSFVSFLQRPARM 228
2615 LIERYGETNG-----SHAVLFSLNYVDFSLF-QLCISLVGKMKLI-I-PPREGLTHEAFY 2668
DB 229 MHLMSDFHAFSAAPNFATELARRTTDDMAGRDIGNILITLISGSEVQAATIKRFADR 288
2669 DIGRREKLSYLSGTPSVLQOIELSR-----LPHLMHVTAAAGEEFHASQFEKMRSD 2718

QY 289 FA 290
DB 2719 FA 2720

RESULT 11
US-08-222-617A-25
Sequence 25, Application US/08222617A
Patent No. 5882879
GENERAL INFORMATION:
APPLICANT: Veenstra, Annemarie E.
APPLICANT: Martin, Juan F.
APPLICANT: Garcia, Bruno D.
APPLICANT: Gutierrez, Santiago
APPLICANT: Barredo, Jose L.
APPLICANT: Von Doehren, Hans
APPLICANT: Palissa, Harriet
APPLICANT: Montenegro, Eduardo P.
TITLE OF INVENTION: A Method for Influencing Beta-Lactam
TITLE OF INVENTION: Antibiotic Production and for Isolation of Large
TITLE OF INVENTION: Quantities of ACV Synthetase
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
STREET: 300 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/222,617A
FILING DATE: 04-APR-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
REFERENCE/DOCKET NUMBER: 97,157
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 3712 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-222-617A-25

Query Match 10.4%; Score 160.5; DB 2; Length 3712;
Best Local Similarity 22.8%; Pred. No. 2.7e-07;
Matches 69; Conservative 55; Mismatches 129; Indels 49; Gaps 11;
DB 7 DGVAI-----TLRSQLYRRTLVNAQELSRGCS--TGDRVISAPOGLEVVAYVYALGALQA 59
2450 DRIADGTRSLSYSELNERANQLVHLIISASIVADDRIALLDKSIDVIALVAMKA 2509
DB 60 GRIAVPLSPQGVTDERSDVSLSPPVAILTSSAVDDVOHVHARRCESPSTIEVD 119
2510 GAAYVPLDPT--YPSQRELIIEESSARTLTTR-----KHPR-----GQIVANVP 2554
DB 120 LLDLDAPN-----GYTFKDEYVST-----AYLYTSGSTRPAGVVMHONVRNFEQ 168
2555 SVYLDSPETLACINQOSKEPPTSTOKPSDLAVITFSGTGKPKGVLVHOSVQLRNS 2614
DB 169 LMSGYADTDGIPPPNSALVSWLPFYHDMGLVIGICAPILIGYPAVLTSFVSFLQRPARM 228
2615 LIERYGETNG-----SHAVLFSLNYVDFSLF-QLCISLVGKMKLI-I-PPREGLTHEAFY 2668
DB 229 MHLMSDFHAFSAAPNFATELARRTTDDMAGRDIGNILITLISGSEVQAATIKRFADR 288
2669 DIGRREKLSYLSGTPSVLQOIELSR-----LPHLMHVTAAAGEEFHASQFEKMRSD 2718

OY 289 FA 290
DB 2719 FA 2720

RESULT 12

US-09-335-409-3
; Sequence 3, Application US/09335409
; Patent No. 6121029
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/335.409
; CURRENT FILING DATE: 1999-06-17
; SOFTWARE: Patentlin Ver. 2.0
; SEQ ID NO 3
; LENGTH: 1410
; TYPE: PRF
; ORGANISM: Sorangium cellulosum
US-09-335-409-3

Query Match

Best Local Similarity 9.4%; Score 144; DB 3; Length 1410;
Matches 72; Conservative 41; Mismatches 111; Indels 60; Gaps 13;

OY 12 TLTSLQRLRLTNAOELSCGSGTGDRLV-ISAPOGLEVVAYLGALGRIAPVLSVPO 70
DB 548 TLTVELSRSRRLGABRLREGARNTLVAVYMEKGEQVAVLVLESGAAYVPL--D 604
OY 71 GGVYDERSDSVLSDSPVALITSSAVD-----VGVHVARPGSGPSIIEV 118
DB 605 ADLPERHAYLL-DHGEVKLVLTQPLDGLKLSMPGLORLVSEAGVEDGDGP----- 658
OY 119 DLDLDAPNGYTFKDEPESTAYLQYTSSTRTPAQVYVMSHO-----NRFVFEQLMS 171
DB 659 -MMPQOTRS-----DLAYVITTSSTGLPKGVAMDHGAVNTIIDIEREE----- 703
OY 172 GYFADTDGIPPNASL-VSMLEPFYHDMGLVIGICAPILGGYPAVLTSFVSFLQRPARMH 230
DB 704 -----IGPDRVLALSLSFSLSYDVFGILA--AGG--TIYVPPASKLRDPAHNAE 751
OY 231 IMASD-FHAFSAAPNPAFELARRTTDDMAGRDILGNITILISG 273
DB 752 LIEREKVYVWNSVPAIMRLVHEFGRPDSLARSIR--LSLSLG 793

RESULT 13

US-08-222-617A-8
; Sequence 8, Application US/08222617A
; Patent No. 5882879
; GENERAL INFORMATION:
; APPLICANT: Veenstra, Annemarie E.
; APPLICANT: Martin, Juan F.
; APPLICANT: Garcia, Bruno D.
; APPLICANT: Gutierrez, Santiago
; APPLICANT: Barredo, Jose L.
; APPLICANT: Von Doehren, Hans
; APPLICANT: Palissat, Harriet
; APPLICANT: Van Liempt, Henk
; APPLICANT: Montenegro, Eduardo P.
; TITLE OF INVENTION: A Method for Influencing Beta-Lactam
; TITLE OF INVENTION: Antibiotic Production and for Isolation of Large
; TITLE OF INVENTION: Quantities of ACV Synthetase
; NUMBER OF SEQUENCES: 27

CORRESPONDENCE ADDRESS:

ADDRESS: McDonnell Boehnen Hulbert & Berghoff
STREET: 300 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentlin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/222,617A
FILING DATE: 04-APR-1994
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

REFERENCE/DOCKET NUMBER: 97,157

INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:
LENGTH: 798 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Bacillus brevis
US-08-222-617A-8

Query Match

Best Local Similarity 9.3%; Score 143; DB 2; Length 798;
Matches 68; Conservative 51; Mismatches 134; Indels 48; Gaps 13;

OY 7 DGVAT-----LTSLQRLRLTNAOELSCGSGTGDRLV-ISAPOGLEVVAYLGALG 60
DB 42 DVAIVAFENRRLSTQELNKAQRLALEKGVQDTSIVGMMEKSIENVIAIILAVLNG 101
OY 61 RIAPVLSV--POGGVYDERSDSVLSDSPVALITSSAVDDVGVHVARPGSGPSIIEV 118
DB 102 GAYVVIDIEYR-----DRIGYIILDQSKIVLTQKS-VSQLV-HDVGYSGE-----VVVL 150
OY 119 DLDLDAPNGYTFKDEPEPS-TAYLQYTSSTRTPAQVYVMSHOVNV--NFEQLMSGFPA 175
DB 151 DEEQDARETANLHQPSTDLAYVITTSSTGLPKGVAMDHGAVNTIIDIEREE----- 205
OY 176 DTDGIPPNASLVMSLP--FYHDMGLVIGICAPILGGYPAVLTSFVSFLQRPARMHLM 232
DB 206 ---LASPSKTSGLPACRSTHPPGKSWLCCLA---PRVHPSKQTHDFAAEHYLS 257
OY 233 ASDFAFSAAPNPAFELARRTTDDMAGRDILGNITILISGSEVQAATIKRFADRFARF 292
DB 258 ENELTITLPPYVTLHPIPERIT-----SLRIMITASASAPLVNWKOKLRXI 307
OY 293 N 293
DB 308 N 308

RESULT 14

US-08-471-119A-2
; Sequence 2, Application US/08471119A
; Patent No. 5827706
; GENERAL INFORMATION:
; APPLICANT: Leitner, Ernst
; APPLICANT: Schergerdoffer, Elisabeth
; APPLICANT: Weber, Gerhard
; TITLE OF INVENTION: Cyclosporin Synthetase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:

```

ADDRESS: No. 5827706artis Corporation
STREET: 59 Route 10
CITY: East Hanover
STATE: New Jersey
COUNTRY: USA
ZIP: 07936

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/471,119A
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Kassenhoff, Melvyn
REGISTRATION NUMBER: 26,389
REFERENCE/DOCKET NUMBER: 100-8029/CONT/CONT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 503 8474
TELEFAX: 201 503 8474
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 15281 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Tolypocladium niveum
STRAIN: ATCC 34921
US-08-471-119A-2

Query Match
Best Local Similarity 9.0%; Score 139; DB 2; Length 15281;
Matches 46; Conservative 30; Mismatches 64; Indels 34; Gaps 5;

QY 7 DGVAI-----TLRSQILRTLNVAQEL-SRCGSGDRVVISAPQGLEVVAVIAGALQAG 60
DB 4570 DSAIALHGSKSLYADLRSDRVARLRLRHSFSSDILLVLAAPRSCETIIAFLGILKAN 4629
QY 61 RIVAPLSVPGGVYTDERSDVSLSDSPVALITLTSSAVDDVYQVHARRPGSPSIIIE--- 117
4630 LAILPLDVKRPA---RIDIVSSLPGNKILILGANV-----TPKRLQDAAI 4673
118 -----VDLIDLDPANGYTFKDEDEPSTAYLYQYSGSTRPAGVYVMSHONV 162
DB 4674 DFPVIRDTFTTLTLDGTLDGPTIERPSAQSLAYAMFTSGTGRPKGVYVQHRNI 4727

RESULT 15
US-08-510-646B-33
; Sequence 33, Application US/08510646B
; Patent No. 6077699
; GENERAL INFORMATION:
; APPLICANT: Blanc, Veronique
; APPLICANT: Blanche, Francis
; APPLICANT: Crouzet, Joel
; APPLICANT: Jacques, Nathalie
; APPLICANT: Lacroix, Patricia
; APPLICANT: Thibaut, Denis
; APPLICANT: Zagorec, Monique
; APPLICANT: Debussche, Laurent
; APPLICANT: De Crey-Lagard, Valerie
; TITLE OF INVENTION: Polypeptides Involved in the
; TITLE OF INVENTION: Biosynthesis of Streptogramins, Nucleotide Sequences
; TITLE OF INVENTION: Coding For These Polypeptides And Their Use
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
```

```

ADDRESS: Finnegan, Henderson, Farabow, Garrett & Dunner
STREET: 1300 I Street, N.W., Suite 700
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/510,646B
FILING DATE: 03-AUG-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/403,852
FILING DATE: 10-MAY-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/FR 93/00923
FILING DATE: 25-SEP-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 92/11441
FILING DATE: 25-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: Meyers, Kenneth J.
REGISTRATION NUMBER: 25,146
REFERENCE/DOCKET NUMBER: 03806, 0054-01000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 33:
SEQUENCE CHARACTERISTICS:
LENGTH: 739 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-510-646B-33

Query Match
Best Local Similarity 8.4%; Score 129; DB 3; Length 739;
Matches 49; Conservative 28; Mismatches 67; Indels 24; Gaps 7;

QY 7 DGVAITLTRSQ-----LYRRTLNVAQELSRCGSGDRVV-ISAPQGLEVVAVIAGALQAG 60
DB 523 DAAVIAVQEGTEGYADLNTFRANLRLRLRNOGIGPEGVVALSPRSADLIVSLAVLAKTS 582
QY 61 RIVAPLSVPGGVYTDERSDVSLSDSPVALITLTSSAVDDVYQVHARRPGSPSIIIEVDL 120
583 AAVLPV---DPAYPARIRIAYLLDQCAPALVLTHTSV-----AAGLPGGVPOLL.-VDO 630
DB 121 LDLDAPNGYTFKDE-----YP-STAYLYQYSGSTRPAGVYVMSHONV 162
DB 631 VGDDVPGHDLTDAERTTPLHLPLPAIVLYTSGTGLPKGVVPHRSV 678
```

Search completed: January 5, 2001, 14:27:33
Job time: 225 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 5, 2001, 14:26:37 : Search time 62.98 Seconds
(without alignments)
162.879 Million cell updates/sec

Title: US-09-461-774-6
Perfect score: 1540
Sequence: 1 DYEQDWGVAILTLRSQLYR.....TIKRFADRFARFNLQERVKA 300

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 268485 seqs, 34193795 residues
al number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: /SIDSL/gcgdata/geneseq/geneseqp/AA1980.DAT:*
2: /SIDSL/gcgdata/geneseq/geneseqp/AA1981.DAT:*
3: /SIDSL/gcgdata/geneseq/geneseqp/AA1982.DAT:*
4: /SIDSL/gcgdata/geneseq/geneseqp/AA1983.DAT:*
5: /SIDSL/gcgdata/geneseq/geneseqp/AA1984.DAT:*
6: /SIDSL/gcgdata/geneseq/geneseqp/AA1985.DAT:*
7: /SIDSL/gcgdata/geneseq/geneseqp/AA1986.DAT:*
8: /SIDSL/gcgdata/geneseq/geneseqp/AA1987.DAT:*
9: /SIDSL/gcgdata/geneseq/geneseqp/AA1988.DAT:*
10: /SIDSL/gcgdata/geneseq/geneseqp/AA1989.DAT:*
11: /SIDSL/gcgdata/geneseq/geneseqp/AA1990.DAT:*
12: /SIDSL/gcgdata/geneseq/geneseqp/AA1991.DAT:*
13: /SIDSL/gcgdata/geneseq/geneseqp/AA1992.DAT:*
14: /SIDSL/gcgdata/geneseq/geneseqp/AA1993.DAT:*
15: /SIDSL/gcgdata/geneseq/geneseqp/AA1994.DAT:*
16: /SIDSL/gcgdata/geneseq/geneseqp/AA1995.DAT:*
17: /SIDSL/gcgdata/geneseq/geneseqp/AA1996.DAT:*
18: /SIDSL/gcgdata/geneseq/geneseqp/AA1997.DAT:*
19: /SIDSL/gcgdata/geneseq/geneseqp/AA1998.DAT:*
20: /SIDSL/gcgdata/geneseq/geneseqp/AA1999.DAT:*
21: /SIDSL/gcgdata/geneseq/geneseqp/AA2000.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	172	11.2	550	17	P. pyralis variant
2	171	11.1	550	16	firefly luciferase
3	170	11.0	550	15	firefly luciferase
4	170	11.0	550	16	firefly luciferase
5	170	11.0	550	20	firefly luciferase
6	170	11.0	743	15	P. pyralis pterLuc
7	170	11.0	743	16	Hepatitis C Virus
8	169	11.0	550	17	Vaccinia virus vec
9	168	10.9	550	18	P. pyralis variant
10	168	10.9	551	18	Modified firefly 1
11	168	10.9	815	13	Lux::npt-ii fusion
12	162	10.5	550	20	P. pyralis luc pro

13	161.5	10.5	3712	12	R13896
14	160.5	10.4	3639	14	R40227
15	150	9.7	563	15	R53923
16	146.5	9.5	545	20	Y06860
17	144.5	9.4	546	20	Y06855
18	144	9.4	1410	21	Y58574
19	143.5	9.3	546	20	Y06859
20	143.5	9.3	546	20	Y06856
21	142.5	9.3	546	20	Y06857
22	139	9.0	15281	15	R44929
23	137	8.9	544	20	Y06862
24	137	8.9	544	20	Y06863
25	136.5	8.9	403	20	W97719
26	136	8.8	543	16	R76569
27	135	8.8	546	20	Y06851
28	135	8.8	1274	14	R34714
29	134.5	8.7	517	21	Y74605
30	134.5	8.7	546	20	Y06858
31	134.5	8.7	546	20	Y06852
32	134.5	8.7	546	20	Y06854
33	134.5	8.7	575	20	Y05559
34	134.5	8.7	3588	14	R34712
35	134	8.7	535	20	Y06092
36	133.5	8.7	546	20	Y06850
37	132.5	8.6	517	21	Y74604
38	132.5	8.6	546	20	Y06853
39	130.5	8.5	552	19	W71636
40	130	8.4	525	20	W99453
41	130	8.4	643	20	Y14943
42	130	8.4	643	20	Y14949
43	128.5	8.3	669	13	R23968
44	128.5	8.3	669	13	R23969
45	126	8.2	3587	14	R34713

ALIGNMENTS

RESULT 1	
R98518	R98518 standard; Protein; 550 AA.
XX	
AC	R98518:
DT	11-MAR-1997 (first entry)
XX	
DE	P. pyralis variant luciferase E270K/E354K.
XX	
KW	Luciferase; firefly; ATP; heat stability; luciferin; light.
XX	
OS	Photinus pyralis.
XX	
FT	Key Location/Qualifiers
FT	Misc-difference 270
FT	/label= E270K
FT	Misc-difference 354
FT	/label= E354K
XX	
PN	W09622376-A1.
XX	
PD	25-JUL-1996.
XX	
PR	19-JAN-1996; 96WO-GB00099.
XX	
PR	24-APR-1995; 95GB-0008301.
XX	
PR	20-JAN-1995; 95GB-0001172.
XX	
PA	(MINA) UK SEC FOR DEFENCE.
XX	
PI	Lowe CR, Murray JA, Squitrell DJ, White PJ;
XX	
DR	WPI; 1996-354533/35.
XX	

ACV synthetase. A
ACVS. Acromonium
Acyl CoA synthetase
Mutant luciferase
Mutant luciferase
Sorangium cellulos
Mutant luciferase
Mutant luciferase
Mutant luciferase
T. niveum Cyclospo
Mutant luciferase
Mutant luciferase
Staphylococcus aur
Beetle luciferase
Mutant luciferase
Bacillus subtilis
Neisseria meningit
Mutant luciferase
Mutant luciferase
Maize 4-coumarate:
Bacillus subtilis
Aspen 4-coumarate
Mutant luciferase
Neisseria meningit
Mutant luciferase
Omega-cyclohexane
B. dinitra pimeyl
Amino acid sequenc
Amino acid product.
faca gene product.
Bacillus subtilis

PT Mutant luciferase(s) with greater heat stability than wild-types -
 PT useful in luminescent reagents

XX
 PS Example 2; Page 26-29; 39pp; English.

CC This sequence represents a variant luciferase from the firefly, P.
 CC pyralis. In this sequence the amino acid at position 270 has
 CC been mutated such that it is not Glu. Also the amino acid at position
 CC 354 has been mutated so that it is not Glu. This luciferase has a
 CC lower Km for the substrate ATP than native luciferase. This
 CC luciferase therefore has a greater heat stability than wild-type
 CC luciferase. The luciferase may be used in an assay to measure ATP
 CC with luciferin, whereby the amount of light is related to the amount
 CC of ATP present.

XX Sequence 550 AA;

Query Match 11.2%; Score 172; DB 17; Length 550;

Best Local Similarity 23.5%; Pred. No. 4.9e-09; Mismatches 118; Indels 68; Gaps 17;

Matches 74; Conservative 55; Mismatches 118; Indels 68; Gaps 17;

DB 9 VATTLRSQLYRRTLVNAQELRSG-STGDRVYISAPQGLEVVAVILGALQAGRIAPLS 67

DB 47 IevnltYaeYfemsvrLaeamkrygIntnhrIvcsensqfImpyIgalIfg-vav--- 102

DB 68 VPOGVTDEERS-DSVLSDDSPVALITTSASVDDVVOHVARRP----- 108

DB 103 aPEndIYneellnsmIsqPvfvskYgIkInVqkIpIqkIlImdsKtdYgfg 162

DB 109 -----GESPSIIEVDLDDAPNGYTFKEDEYPTAVLQYTSSTRTPGVNVSHON 161

DB 163 smYtfvshlpPgfnedyf---pesf---drkktIalImssgStgIpKvAlphrt 214

DB 162 VRNFEQLMSGYRADPDGIPPNLSALVSWLPFYHDMGL-VIG--ICAPILGYPALVLS 217

DB 215 acrrfshardpIfgnq---ItpdtaIIsVvrfhgfmftIlgYlIc-----gfrvvl-- 264

DB 218 PVSFLOPRAPRMHMLA-SDFHAFSA--APNFAFELAAARTTDDMAGRDGLNLTLLSG- 273

DB 265 ----myrfekelfIrlsIdgYkIqsalIvptl-fsfakstIdky---dIsnlheIasg 316

DB 274 ---SERVOAATIKRF 285

DB 317 apIskevgeavakrf 331

DB 317 apIskevgeavakrf 331

DB 317 apIskevgeavakrf 331

DB 317 apIskevgeavakrf 331

DB 317 apIskevgeavakrf 331

DB 317 apIskevgeavakrf 331

DB 317 apIskevgeavakrf 331

DB 317 apIskevgeavakrf 331

DB 317 apIskevgeavakrf 331

DB 317 apIskevgeavakrf 331

DB 317 apIskevgeavakrf 331

DB 317 apIskevgeavakrf 331

DB 317 apIskevgeavakrf 331

DB 317 apIskevgeavakrf 331

DB 317 apIskevgeavakrf 331

PD 28-SEP-1995.

XX 22-MAR-1995; 95WO-GB00629.

XX 20-JAN-1995; 95GB-0001170.

XX 23-MAR-1994; 94GB-0005750.

XX (MINA) UK SEC FOR DEFENCE.

XX Lowe CR, Murray JAH, Squirrel DJ, White PJ;

XX WPI; 1995-344619/44.

XX New heat stable luciferase mutants - related DNA, vectors and for

XX PT transformed cells, useful as reporter genes, assay labels and for

XX PT determ of ATP

XX A doubly mutated luciferase of Photinus pyralis has leucine at

XX amino acid position 215 and an amino acid other than glutamic

XX acid at position 354. The luciferase, obtd. by site-directed

XX mutagenesis of the luc gene, shows improved thermostability

XX compared to the wild-type.

XX Sequence 550 AA;

Query Match 11.1%; Score 171; DB 16; Length 550;

Best Local Similarity 23.5%; Pred. No. 6.2e-09; Mismatches 118; Indels 68; Gaps 17;

Matches 74; Conservative 55; Mismatches 118; Indels 68; Gaps 17;

DB 9 VATTLRSQLYRRTLVNAQELRSG-STGDRVYISAPQGLEVVAVILGALQAGRIAPLS 67,

DB 47 IevnltYaeYfemsvrLaeamkrygIntnhrIvcsensqfImpyIgalIfg-vav--- 102

DB 68 VPOGVTDEERS-DSVLSDDSPVALITTSASVDDVVOHVARRP----- 108

DB 103 aPEndIYneellnsmIsqPvfvskYgIkInVqkIpIqkIlImdsKtdYgfg 162

DB 109 -----GESPSIIEVDLDDAPNGYTFKEDEYPTAVLQYTSSTRTPGVNVSHON 161

DB 163 smYtfvshlpPgfnedyf---pesf---drkktIalImssgStgIpKvAlphrt 214

DB 162 VRNFEQLMSGYRADPDGIPPNLSALVSWLPFYHDMGL-VIG--ICAPILGYPALVLS 217

DB 215 acrrfshardpIfgnq---ItpdtaIIsVvrfhgfmftIlgYlIc-----gfrvvl-- 264

DB 218 PVSFLOPRAPRMHMLA-SDFHAFSA--APNFAFELAAARTTDDMAGRDGLNLTLLSG- 273

DB 265 ----myrfekelfIrlsIdgYkIqsalIvptl-fsfakstIdky---dIsnlheIasg 316

DB 274 ---SERVOAATIKRF 285

DB 317 apIskevgeavakrf 331

DB 317 apIskevgeavakrf 331

DB 317 apIskevgeavakrf 331

DB 317 apIskevgeavakrf 331

DB 317 apIskevgeavakrf 331

DB 317 apIskevgeavakrf 331

DB 317 apIskevgeavakrf 331

DB 317 apIskevgeavakrf 331


```

XX AC Y08523;
XX DT 03-AUG-1999 (first entry)
XX DE P. pyralis pTeluclt Luciferase protein.
XX KM Firefly; luciferase; tetracycline; transcriptional control; TetR; TetA;
XX KM tetracycline repressor; tetracycline promoter; luminescence; luxCDABE;
XX KM insect; Tn10; medicine; dosage; cheese production; antibiotic; foodstuff;
XX KM allergy.
XX OS Photinus pyralis.
XX PN W09925866-A1.
XX PD 27-MAY-1999.
XX PF 11-NOV-1998; 98MO-F100873.
XX 14-NOV-1997; 97FI-0004235.
PA (KARP/) KARP M.
PA (KORP/) Korpela M.
PA (KUR/) Kurittu J.
PI Karp M, Korpela M, Kurittu J;
PI WPI: 1999-338015/28.
DR N-PSDB; V72416.
PT Assaying for tetracycline using recombinant prokaryotic cells
XX Example 2; Page 25-26; 67pp; English.
XX CC This invention describes a novel tetracycline assay that uses recombinant
XX CC prokaryotic cells comprising a luciferase gene under the transcriptional
XX CC control of a tetracycline repressor and tetracycline promoter and
XX CC involves the detection of luminescence emitted from the cells. The assay
XX CC can be used to distinguish tetracycline form other microbial agents. The
XX CC invention also describes a novel plasmid comprising either the luxCDABE
XX CC genes, a tetracycline repressor (TetR) and a tetracycline promoter (TetA)
XX CC from Tn10, or the insect luciferase gene, a tetracycline repressor (TetR)
XX CC and a tetracycline promoter (TetA) from Tn10. The tetracycline assay
XX CC method can be used for the determination of tetracycline in a sample,
XX CC e.g. to study the dosage and penetration of the medicine. In a sample,
XX CC also be used to test cheese production, as cheese making bacteria are not
XX CC able to work in the presence of tetracycline. The method can also be used
XX CC to determine the presence or concentration of antibiotics in foodstuffs,
XX CC e.g. for allergic people. The present assay method does not rely on the
XX CC growth of microbes as do conventional tests, and so is much more rapid.
XX CC The present assay is also more sensitive, as even a small amount of
XX CC luminescence can be detected.
XX Sequence 550 AA;
XX
XX Query Match 11.0%; Score 170; DB 20; Length 550;
XX Best Local Similarity 23.5%; Pred. No. 7.9e-09;
XX Matches 74; Conservative 54; Mismatches 119; Indels 68; Gaps 17;
OY 9 VAITLRSQLYRRLTVAQELSRG-STGDRVISAPOGLEVVAVYALGALQAGRIAPLS 67
OY 47 Ievnltyaeyfemsyvlaeamkryglntnhrlyvcsensqfmpylgalfig-vav--- 102
OY 68 VPQGGVTDERS-DSVLSDSPVAIILTSSAIVDDVOHVARRP----- 108
OY 103 apandynellnsmisqptvfvskgllqkllnvqkklpikllindsktdyqfg 162
OY 109 -----GESPSITIEVDLIDLAPNGYTFKEDYESTAYLQYTSSTRTTRPACVVMHON 161
OY 163 smytftslhlpqpfneydiv-----pesf-----drdktrialmssgstglpkxvalphrt 214

```

```

OY 162 VRVNFEOIMSGYFADDDGIPPNASALVSWLPFYHDMGL--VIG--ICAPILGIPAVLTS 217
OY 215 acvrfshardpifgnq---lptdailsvpfnhgfmgfcllyllc-----gfrvll-- 264
OY 218 PVSFLQRPARMHMLMA-SDHFHFA--APNFAFELARRTDDDMGRDGNILTLTSG- 273
OY 265 ---myrfeeelflrsldykglsallvpcl-tsflakslclcky---disnlhelasg 316
OY 274 ---SERVOATIKRF 285
OY 317 aplskeygeavakrf 331
Db
RESULT 6
R54867
ID R54867 standard; Protein; 743 AA.
XX AC R54867;
XX DT 21-DEC-1994 (first entry)
XX DE Hepatitis C Virus core protein-luciferase fusion protein.
XX KM Hepatitis C Virus; HCV; control; Non-A, non-B hepatitis virus;
XX KM antisense; therapy; inhibition; viral protein precursor;
XX KM recombinant vaccinia virus; HCV core protein gene; firefly;
XX KM luciferase reporter gene; fusion construct.
XX OS Chimeric Hepatitis C virus.
XX OS Chimeric Photinus pyralis.
XX PN CA2104649-A.
XX PD 26-FEB-1994.
XX PF 23-AUG-1993; 93CA-2104649.
XX PR 25-AUG-1992; 92JP-0248796.
XX PR 03-MAR-1993; 93JP-0042736.
XX PA (SEKI/) SEKI M.
XX PI Honda Y, Seki M, Yamada E;
XX DR WPI: 1994-151836/19.
XX PT N-PSDB; Q65322.
XX PT Anti-sense oligo:nucleotide(s) complementary to the hepatitis C
XX PT virus genome - are useful as antiviral agents
XX PS Example 5; Page 244-250; 262pp; English.
XX CC A recombinant vaccinia virus which codes for a HCV core protein
XX CC fused to the firefly luciferase enzyme was constructed from PCR
XX CC amplified fragments. The construct (which encodes the fusion
XX CC protein sequence R54867) was useful for assaying the inhibitory
XX CC activity of various antisense oligonucleotides on HCV gene
XX CC translation; detection of luciferase activity provided a measure
XX CC of expression of the HCV core protein.
XX Sequence 743 AA;
XX
XX Query Match 11.0%; Score 170; DB 15; Length 743;
XX Best Local Similarity 23.5%; Pred. No. 1.2e-08;
XX Matches 74; Conservative 54; Mismatches 119; Indels 68; Gaps 17;
OY 9 VAITLRSQLYRRLTVAQELSRG-STGDRVISAPOGLEVVAVYALGALQAGRIAPLS 67
OY 240 Ievnltyaeyfemsyvlaeamkryglntnhrlyvcsensqfmpylgalfig-vav--- 295
OY 68 VPQGGVTDERS-DSVLSDSPVAIILTSSAIVDDVOHVARRP----- 108

```


Sequence	550 AA;
Query Match	11.0%; Score 169; DB 17; Length 550;
Best Local Similarity	23.5%; Pred. No. 9.9e-09;
Matches 74; Conservative 54; Mismatches 119; Indels 68; Gaps	
QY	9 VAITITRSQLYKRTILNVAQELSRGSGTGDVIVISAPOGLEFYVAIVLGAQAGIINPLS 67
DB	47 IevnltlyaeYfemsyrlaeamkryglfnhrliwconsliqfimpvlgalfig-vav--- 1020
QY	68 VPQGVGVTEBRS-DSVLSQSSPVAILITTSAAVDVQVHARRP----- 1080
DB	103 apandiylereellsmnislqptvrvskkylqkllnvqkllpiklllmsdktcygfg 1620
QY	109 -----GESPSIIIEVLLDLDAPNGTFEDEXPSTAYLDQYSGSTRPAGVMSQN 1610
DB	163 smytlvtshlppgfyeyfv-----pest-----dtktlalmsgsctgtpkgyalphrt 2140
QY	162 VRVNEFQLMGSGPADFDGIPPNNSALVSWLPFYHDMGL-VIG--ICAPILGYPVAVLTS 2170
DB	215 acvrfshardpifgq---ltpdtaislsvpfnhgfgmfttlygllic-----gfrvll-- 2640
QY	218 PVSEIQRAPRRMMHMA-SDFHAFSA--APNPAFELARTDDDDMAGROGLNLTILSG- 2730
DB	265 ---myrflexelfirslqdykxigallvpl-lfsafakstldiky---dlsnlhelasgg 3160
QY	274 ---SERVOAATIKRP 285
DB	317 aplskvgeavakrf 331
RESULT 9	
W32936	
ID	W32936 standard; Protein; 550 AA.
XX	W32936;
AC	W32936;
XX	
XX	22-JAN-1998 (first entry)
DT	
XX	Modified firefly luciferase luc+.
DE	
XX	Firefly; luciferase; mutant; reporter gene; cell culture;
KM	transgenic plant; transgenic animal; cell free expression system;
KM	peroxisomal translocation sequence; cytoplasmic form; luc+;
KM	restriction site; genetic regulatory site; N-glycosylation site;
KM	post-translational modification.
Phoctinus pyralis.	
Synthetic.	
XX	
FH	Key Location/Qualifiers
FT	Misc-difference 50
FT	Misc-difference /note= "wild type Asn replaced by Asp"
FT	Misc-difference 119
FT	Misc-difference /note= "wild type Asn replaced by Gly"
FT	Misc-difference 548
FT	Misc-difference /note= "wild type Ser replaced by Ile"
FT	Misc-difference 549
FT	Misc-difference /note= "wild type Lys replaced by Ala"
FT	Misc-difference 550
FT	/note= "wild type Lys replaced by Val"
XX	
XX	US5670356-A.
PD	
XX	23-SEP-1997.
XX	
XX	12-DEC-1994; 94US-0354240.
XX	
XX	12-DEC-1994; 94US-0354240.
XX	
PA	(PROM-) PROMEGA CORP.
XX	

Pt	Sherff BA; Wood KV;
Dx	WPI; 1997-479459/44.
Dr	N-PsDB; T88243.
Xx	
Pf	Modified Luciferase gene - encoding cytoplasmic form of luciferase enzyme
Pf	
Xx	
P5	Claim 5; Columns 25-28; 31pp; English.
Cc	The present sequence is encoded by the modified firefly luciferase cDNA luct+, which can be used as a reporter gene in experimental biological systems, e.g. cell cultures, transgenic plants and animals, and in cell free expression systems. Removal of the cDNA's peroxisomal translocation sequence yields a cytoplasmic form of the enzyme with optimal substrate availability. Potentially interfering restriction sites and genetic regulatory sites are removed, and codon usage for mammalian cells is improved. Absence of potential N-glycosylation sites minimises post-translational modification.
Cc	
SQ	Sequence 550 AA;
Oy	Query Match 10.9%; Score 168; DB 18; Length 550; Best Local Similarity 23.5%; Pred. No. 1.3e+08; Matches 74; Conservative 53; Mismatches 120; Indels 68; Gaps 17.
Dd	9 VAITLRSGLYRRLTNVAQELSRGC-STGDRVVISAPOGLEVVAVVLGAGRIAPLPS 67 : 47 Ievdltyaeylemsvrlaeamkryglnhtrlrvcsenslgfmpylgalflg-vav--- 102 :
Oy	68 VPQGVTPEERS-DSVLSDSPVALITTSAAVDVVQHVARRP----- 108 Db apendilnerellnsmgsqplevfvsakqlqklinvqkkpldqklmdsktdygfq 162 :
Oy	109 -----GESPPSTIEVDLLIDDAENGRTFKDEPERSTAYLOYTSGSTRPAGVMSHON 161 : 163 smytftvtshllpglfeydfv----peef-----drcktlalimssqtglpkgyvalphrt 214 :
Oy	162 VRVNFEQLMGSVFATPDGIIPRNASLVSLPRYHMIGL-VVG--ICAPTLCGYPAVLNS 217 Db acvrfsnarqpfifngq---llpdctailsvrbhhfgmftltlgyllic-----qflravl-- 264 :
Oy	218 PVSFQRPARRMHLMA-SDFNAFSA--APNFAFELAARTRTTDDDMAGRDIGNILTLISG- 273 Db 265 ----myrfteeeelflsldqykigsallpytll-fsfrafkstlldky---dlshlhelaagsg 316 : Oy 274 ---SERVOATIKRF 285 : Db 317 aplskevgavaakrf 331 :
Result	10
ID	w32937
AC	w32937 standard; Protein; 551 AA.
XX	
WT	w32937;
XX	
DT	22-JAN-1998 (first entry)
XX	
DE	Modified firefly luciferase luc+NF.
XX	
KW	Firefly; luciferase; mutant; reporter gene; cell culture;
KW	transgenic plant; transgenic animal; cell free expression system;
KW	peroxisomal translocation sequence; cytoplasmic form; luc+NF;
KW	restriction site; genetic regulatory site; N-glycosylation site;
KW	post-translational modification.
XX	
OS	Photinus pyralis.
XX	
Key	Location/Qualifiers

```
FT Misc-difference 2..3 /note= "wild type Glu replaced by Val-Thr"
FT Misc-difference 51 /note= "wild type Asn replaced by Asp"
FT Misc-difference 120 /note= "wild type Asn replaced by Gly"
FT Misc-difference 549 /note= "wild type Ser replaced by Ile"
FT Misc-difference 550 /note= "wild type Lys replaced by Ala"
FT Misc-difference 551 /note= "wild type Lys replaced by Val"
FT US5670356-A.
PN 23-SEP-1997.
XX 12-DEC-1994; 94US-0354240.
XX 12-DEC-1994; 94US-0354240.
XX (PROM-) PROMEGA CORP.
XX Sheriff BA, Wood KV;
XX MPI: 1997-479459/44.
XX N-PSDB; T88244.
XX Modified luciferase gene - encoding cytoplasmic form of luciferase
XX enzyme
XX Disclosure: Columns 31-36; 31pp; English.
XX The present sequence is encoded by the modified firefly luciferase
XX cDNA luc+NF, which can be used as a reporter gene in experimental
XX biological systems, e.g. cell cultures, transgenic plants and
XX animals, and in cell free expression systems. Removal of the cDNA's
XX peroxisomal translocation sequence yields a cytoplasmic form of
XX the enzyme with optimal substrate availability. Potentially
XX interfering restriction sites and genetic regulatory sites are
XX removed, and codon usage for mammalian cells is improved. Absence
XX of potential N-glycosylation sites minimises post-translational
XX modification. The variation at 5'-end of the cDNA was designed
XX for easier construction of amino-terminal fusions.
XX Sequence 551 AA:
SQ
Query Match 10.9%; Score 168; DB 18; Length 551;
Best Local Similarity 23.5%; Pred. No. 1.3e-08;
Matches 74; Conservative 53; Mismatches 120; Indels 68; Gaps 17;
OY 9 VAITTLRSQLYRRTLNVAOELSRG-STGDRVVISAPOGLEYVAVYALQAGRIAPLVS 67
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 48 Ievdltiaeyfemsrvlaeamkrqylntnhrivcsenslqfmpylgalfg-vav--- 103
OY 68 VPQGGVTDERS-DSVLSDDSPAIIITSSAVDDVQHVARRP----- 108
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 104 apandlynereilmsnglsqrvlvfvskkqlqkllnvqkklplqkllmsktdyqfg 163
OY 109 -----GESPSIIIEVDLLDAPNGYTFKEDEYSTAVLYOTSGSTRTPAQVMSHON 161
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 164 smytlvtshlppgfneydiv---pef-----drkctialimssgsglpxyvalphrt 215
OY 162 VRVNFEOILMSGYFADTDGIPPNASALVSLPFYHDMGL--VIG--ICAPILGYPVAVLNS 217
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 216 acrcfshardplfgng---ltpotalisvvrflhgmfttlyllc-----gfrvvl-- 265
OY 218 PVSEIOPARPMHMLMA-SDFHAFSA--APNFAELAAARTTDDMAGRLGNILTLTSG- 273
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 266 ---myrfeeeelfrlsldgkqisallvptl-fsffakstlldky---dlsmhelaasg 317
OY 274 ---SERVOAATIKRF 285
```

```
DB 318 aplskvegeavakrf 332
RESULT 11
R28127
ID R28127 standard; Protein; 815 AA.
XX R28127:
XX 12-MAR-1993 (first entry)
XX Lux::npt-II fusion protein.
XX Neomycin phosphotransferase-II; genetic selection; genetic marker;
XX gene recovery.
XX Synthetic.
XX WO9217593-A.
XX 15-OCT-1992.
XX 30-MAR-1992; 92WO-CA00139.
XX 28-MAR-1991; 91US-0676432.
XX (CANA ) NAT RES COUNCIL CANADA.
XX Crosby WL, Datla RSS, Hammerlindl JK, Selvaraj G;
XX MPI: 1992-366262/44.
XX N-PSDB; Q30000.
XX Fused gene with characteristics of component parts - comprising a
XX gene conferring conditional growth advantage and a marker gene,
XX used as a probe to select and isolate genetic elements
XX Disclosure: Page 33; 53pp; English.
XX The sequence is that of the fusion protein encoded by the firefly
XX (P.pyralis) luciferase:neomycin phosphotransferase-II chimeric gene.
XX See also R28125.
XX Sequence 815 AA:
SQ
Query Match 10.9%; Score 168; DB 13; Length 815;
Best Local Similarity 23.5%; Pred. No. 2.3e-08;
Matches 74; Conservative 53; Mismatches 120; Indels 68; Gaps 17;
OY 9 VAITTLRSQLYRRTLNVAOELSRG-STGDRVVISAPOGLEYVAVYALQAGRIAPLVS 67
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 47 Ievnltiaeyfemsrvlaeamkrqylntnhrivcsenslqfmpylgalfg-vav--- 102
OY 68 VPQGGVTDERSDV-LSDDSPAIIITSSAVDDVQHVARRP----- 108
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 103 apandlynereilmsnglsqrvlvfvskkqlqkllnvqkklplqkllmsktdyqfg 162
OY 109 -----GESPSIIIEVDLLDAPNGYTFKEDEYSTAVLYOTSGSTRTPAQVMSHON 161
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 163 smytlvtshlppgfneydiv---pef-----drkctialimssgsglpxyvalphrt 214
OY 162 VRVNFEOILMSGYFADTDGIPPNASALVSLPFYHDMGL--VIG--ICAPILGYPVAVLNS 217
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 215 acrcfshardplfgng---ltpotalisvvrflhgmfttlyllc-----gfrvvl-- 264
OY 218 PVSEIOPARPMHMLMA-SDFHAFSA--APNFAELAAARTTDDMAGRLGNILTLTSG- 273
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 265 ---myrfeeeelfrlsldgkqisallvptl-fsffakstlldky---dlsmhelaasg 316
OY 274 ---SERVOAATIKRF 285
```

Db 317 apiskevgeavakrf 331

RESULT 12

ID W92747 standard; Protein; 550 AA.

AC W92747;

DT 06-MAY-1999 (first entry)

XX P. pyralis luc protein.

XX luc protein; luciferase; Km; mutant; steady-state ATP concentration;

XX light intensity.

OS Photinus pyralis.

XX W09846729-A2.

XX 22-OCT-1998.

PF 07-APR-1998; 98WO-GB01026.

XX 11-APR-1997; 97GB-0007486.

XX (MINA) UK SEC FOR DEFENCE.

XX Lowe CR, Murray JAH, Squirrel DJ, White PJ;

XX WPI; 1999-080738/07.

PT Mutant luciferase with increased Km for ATP - useful for measuring

XX the ATP concentration in a cell

XX Disclosure; Fig 8; 30pp; English.

XX This invention describes a novel recombinant mutant luciferase having a

XX Km for ATP which is increased as compared to the wild type luciferase.

XX This mutant can be used to measure the steady-state ATP concentration of

XX a biological sample, e.g. a whole cell. Only when the Km is in the same

XX order of magnitude or greater than the ATP concentration, can there be a

XX degree of proportionality between changes in ATP concentration and

XX changes in light intensity emitted by the luciferase. This sequence

XX represents a Photinus pyralis luc protein used in the method of the

XX invention.

XX Sequence 550 AA;

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

OY 274 ---SERVOATIKRF 285
|: | | |
Db 317 apiskevgeavakrf 331

RESULT 13

ID R13896 standard; Protein; 3712 AA.

XX R13896;

AC R13896;

DT 22-NOV-1991 (first entry)

XX ACV synthetase.

XX Beta lactam antibiotics; penicillin.

XX Acremonium chrysogenum.

XX Key

XX Location/Qualifiers

XX 301..1068

XX /label= I

XX /function= activation of amino acid substrate

XX 374..423

XX /label= subdomain

XX 474..501

XX /label= subdomain

XX 655..699

XX /label= subdomain

XX 725..754

XX /label= subdomain

XX 1392..2154

XX /label= II

XX /function= activation of amino acid substrate

XX 1470..1518

XX /label= subdomain

XX 1564..1590

XX /label= subdomain

XX 1745..1789

XX /label= subdomain

XX 1817..1846

XX /label= subdomain

XX 2474..3295

XX /label= III

XX /function= activation of amino acid substrate

XX 2554..2603

XX /label= subdomain

XX 2647..2673

XX /label= subdomain

XX 2827..2871

XX /label= subdomain

XX 2899..2928

XX /label= subdomain

XX 3560..3647

XX /label= IV

XX /function= thioesterase

XX EP445868-A.

XX 11-SEP-1991.

XX 27-FEB-1991;

XX 91EP-0200423.

XX 27-FEB-1991;

XX 91EP-0200423.

XX 28-FEB-1990;

XX 90EP-0200475.

XX 28-FEB-1990;

XX 90EP-0200486.

XX 02-JUL-1990;

XX 90EP-0201768.

XX 03-OCT-1990;

XX 90EP-0202628.

XX (KONN) GIST-BROCADES NV.

XX Veenstra AE, Martin JF, Garcia BD, Gutierrez S, Barredo JL;

PI

PI	Montenegro PE, Von Döhren H, Palissa H, Van Llempt H;
xx	
DR	WPI; 1991-268735/37.
xx	
DR	N-PSDB; Q13608.
xx	
PT	DNA encoding amino:adipyl-L-cysteinyl-valine synthetase - used for
PT	prodn. of the enzyme or enhanced prodn. of new or known
PT	beta-lactam antibiotic cpds.
xx	
PS	Claim 1; Page 20; 54pp; English.
xx	
CC	The DNA sequence was obt'd. from five subclones isolated from a
CC	gene library of A. chrysogenum C10 (ATCC 48). The protein of
CC	gene library of A. chrysogenum C10 (ATCC 48). The protein of
CC	homology have been deduced from the DNA. Three distinct regions of
CC	these domains several even more conserved elements can be
CC	distinguished. Since the enzyme synthesizes a tripeptide, which
CC	role for these domains in the activation reactions seems likely.
CC	A fourth domain is thought to act as a thioesterase.
CC	The gene can be used to express the synthetase enzyme which can
CC	be used for the prodn. of new beta-lactam antibiotics.
CC	See also R13896.
SQ	Sequence 3712 AA;
xx	
xx	
Query Match	10.5%; Score 161.5; DB 12; Length 3712;
Best Local Similarity	22.8%; Pred. No. 9,8e-07;
Matches 69; Conservative 55; Mismatches 129; Indels 49; Gaps 11	
OY	7 DGVAI-----TLTRSQLVRRRLTNVAOELSRGGS--TGRVVTASAPGLEVVAYALCALQA 59
Db	2450 dtrialdgrrlsyselneranglvhlissasiavddtrialldksidmvalialawwka 2509
OY	60 GRIAAPLSPGGCVTDKERSDVLSSSPVALITTSADVVDVQHVARRGESPPSILEVD 119
Db	2510 gaayvpdppt---ypsqrtellileessartilttr-----khprr---ggtvanvp 2554
OY	120 LLDLDAPN-----GYTFKDEYRPT-----AVLOYSGSTRTPAGVSHONVRNFEQ 168
Db	2555 avdlspellelaingqskentstcqpksdlayivftsgctckpkpgvlyehgsavgjlns 2614
OY	169 LMSGVFADTDGIIPPNSALISWLPEYHDMGLVIGICAPILGSGYPVALTSPVSFLORPARW 228
Db	2615 llyeryfgetng----shavlflsnvyfdtsle-qtlcslvggnklil-preegltheaty 2668
OY	229 MHLMASDFIAFSAAPNFAFELAAARTTTDDMAGRDLGNLTLLTSGSERVOAATIKFFADR 288
Db	2669 digirrekislystpsviqltelr-----lphlmvtlaageethaefekmrsq 2718
OY	289 FA 290
Db	2719 fa 2720
RESULT 14	
R40227	
xx	R40227 standard; Protein: 3639 AA.
xx	
AC	R40227:
xx	
DT	21-FEB-1994 (first entry)
xx	
DE	ACVS.
xx	
KW	Delta-(L-alpha-aminoacidipyl)-L-cystinyl-D-valine synthase; ACVS;
xx	beta-lactam; antibiotic; transformed; cephalosporin; vector.
OS	Acremonium chrysogenum.
xx	
JM	JP05192162-A.
xx	

```

PD      03-AUG-1993.
XX XX
PF      25-JUL-1991;          91JP-0186222.
XX XX
PR      31-JUL-1990;          90JP-0205677.
XX XX
PA      (TAKE ) TAKEBA CHEM IND LTD.
XX XX
DR      WPI; 1993-277475/35.
XX XX
N-PSDB; Q48231.
XX XX
PT      DNA coding delta-(L-alpha-amino-adipyl) L-cystinyl D-valine
PT synthase - for improved productivity of cephalosporin antibiotics
XX XX
PS      Claim 1; Page 14-27; 69pp; Japanese.
XX XX
CC      The sequence (Q48231) is of a vector which includes the
CC delta-(L-alpha-aminoadipyl)-L-cystinyl-D-valine synthase gene.
CC This sequence was transformed into a host cell to express the ACVS
CC product. The protein produced (R40227) was then used to manufacture
CC a beta-lactam antibiotic.
XX XX
SQ      Sequence      3639 AA;

Query Match              10.4%; Score 160.5; DB 14; Length 3639;
Best Local Similarity    22.8%; Pred. No. 1.2e-06;
Matches   69; Conservative 55; Mismatches 129; Indels 49; Gaps 11.

OY      7 DGVAI-----TLTRSQLRTILNAOEISRCGS--TGDRIVISADPGLGVVAYIGALQA 59
Db      2377 driaiaadgtrslyselneranqlvhlillsasiavaddriaillldksidmviaiallawka 2436
OY      60 GRIAEPLSVPGGCVDERSDSVLSPSSVALITTSASAVDDVVQHARRRGESPSEIEVD 119
Db      2437 gaaypllpbt---ypsqrtellleessarclltc-----khcpr-----qgitvanvp 2481
OY      120 LLDDAPN----GYTFEDEEXPTST-----AYLOYTSGSTRTPAGVVMASHONRVNFEO 168
Db      2482 svidaspctlaclnqskpenptstcqkpsdiayivftsgtfgkpxgvlvbhgsvqvlns 2541
OY      169 LMSGCFADTDGIPPNSALVSWLPFYHMGVLGICAPLIGYPRAVLTSPPVSFLORPARW 228
Db      2542 llerfygtngt---shavilfnisyvfafsl-qiclsavgnklll-preegltheafy 2595
OY      229 MHLMASDHAFSAAPNFAPELLAARFTTODDMAGRDLGNLTILSSSERVOATIRFADR 288
Db      2596 digtreklisyslgtpsvlqgielstf-----lphmvtaageefhasqtekmrsq 2645
OY      289 FA 290
Db      2646 fa 2647

RESULT 15
ID      R53923 standard; Protein; 563 AA.
XX AC      R53923;
XX DT      02-FEB-1995 (first entry)
XX DE      Acyl CoA synthetase.
XX KM      Acyl Co-A synthetase; P. fragi; transformation; E. coli.
XX OS      Pseudomonas fragi.
XX PN      JRP06090747-A.
XX PP      05-APR-1994.
XX PF      08-SEP-1992;          92JP-0239842.

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 12, 2001, 18:01:37 ; Search time 3358.72 Seconds
(without alignments)
901.306 Million cell updates/sec

Title: US-09-461-774-7

Perfect score: 432

Sequence: 1 gccacacacccctcccttca.....agatccggtccaccaactga 432

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 7991742 segs, 3503743658 residues

total number of hits satisfying chosen parameters: 15983484

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: em_est41:*
42: em_estfun:*
43: em_esthum1:*

44: em_esthum2:*
45: em_esthum3:*
46: em_esthum4:*
47: em_esthum5:*
48: em_esthum6:*
49: em_esthum7:*
50: em_esthum8:*
51: em_esthum9:*
52: em_esthum10:*
53: em_esthum11:*
54: em_esthum12:*
55: em_esthum13:*
56: em_esthum14:*
57: em_esthum15:*
58: em_esthum16:*
59: em_esthum17:*
60: em_esthum18:*
61: em_esthum19:*
62: em_esthum20:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estov1:*
68: em_estov2:*
69: em_estp11:*
70: em_estp12:*
71: em_estp13:*
72: em_estp14:*
73: em_estp15:*
74: em_estro1:*
75: em_estro2:*
76: em_estro3:*
77: em_estro4:*
78: em_estro5:*
79: em_estro6:*
80: em_estro7:*
81: em_estro8:*
82: em_estro9:*
83: em_estro10:*
84: em_estro11:*
85: em_estro12:*
86: em_estro13:*
87: gb_est41:*
88: gb_est42:*
89: gb_est43:*
90: gb_est44:*
91: gb_est45:*
92: gb_est46:*
93: gb_est47:*
94: gb_est48:*
95: gb_est49:*
96: gb_est50:*
97: gb_est51:*
98: gb_est52:*
99: gb_est53:*
100: gb_est54:*
101: gb_est55:*
102: gb_est56:*
103: gb_est57:*
104: gb_est58:*
105: gb_est59:*
106: gb_est60:*
107: gb_est61:*
108: gb_est62:*
109: gb_est63:*
110: gb_est64:*
111: gb_est65:*
112: em_esthum21:*
113: em_esthum22:*
114: em_esthum23:*
115: em_estom1:*
116: em_estom2:*

117: em_estp16:*
 118: em_estp17:*
 119: em_estp18:*
 120: em_estp19:*
 121: em_estp20:*
 122: em_estp21:*
 123: em_estp22:*
 124: em_estp23:*
 125: em_estp24:*
 126: em_estp25:*
 127: em_estp26:*
 128: em_estp27:*
 129: em_estp28:*
 130: em_estp29:*
 131: em_estp30:*
 132: em_estp31:*
 133: em_estp32:*
 134: em_estp33:*
 135: em_estp34:*
 136: em_estp35:*
 137: em_estp36:*
 138: em_estp37:*
 139: em_estp38:*
 140: em_estp39:*
 141: em_estp40:*
 142: em_estp41:*
 143: em_estp42:*
 144: em_estp43:*
 145: em_estp44:*
 146: em_estp45:*
 147: em_estp46:*
 148: em_estp47:*
 149: em_estp48:*
 150: em_estp49:*
 151: em_estp50:*
 152: em_estp51:*
 153: em_estp52:*
 154: em_estp53:*
 155: em_estp54:*
 156: em_estp55:*
 157: em_estp56:*
 158: em_estp57:*
 159: em_estp58:*
 160: em_estp59:*
 161: em_estp60:*
 162: em_estp61:*
 163: em_estp62:*
 164: em_estp63:*
 165: em_estp64:*
 166: em_estp65:*
 167: em_estp66:*
 168: em_estp67:*
 169: em_estp68:*
 170: em_estp69:*
 171: em_estp70:*
 172: em_estp71:*
 173: em_estp72:*
 174: em_estp73:*
 175: em_estp74:*
 176: em_estp75:*
 177: em_estp76:*
 178: em_estp77:*
 179: em_estp78:*
 180: em_estp79:*
 181: em_estp80:*
 182: em_estp81:*
 183: em_estp82:*
 184: em_estp83:*
 185: em_estp84:*
 186: em_estp85:*
 187: em_estp86:*
 188: em_estp87:*
 189: em_estp88:*

190: gb_gss25:*
 191: gb_gss26:*
 192: gb_gss27:*
 193: gb_gss28:*

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	94.4	21.9	120	7	AA465078	AA465078 EST_Lubert
2	50.6	11.7	632	105	BE214663	BE214663 HV_CEB000
3	49.4	11.4	515	36	AV434394	AV434394 AV434394
4	47.2	10.9	925	190	CNS0091P	AL053013 Drosophila
5	45.8	10.6	925	190	CNS0091P	AL053013 Drosophila
6	42.6	9.9	414	29	AU093617	AU093617 A093617
7	41.8	9.7	344	88	AM287529	AM287529 IGI_242-C
8	41.8	9.7	389	110	BE593608	BE593608 WSI_98-GO
9	41.6	9.6	531	109	BE552631	BE552631 946082H01
10	41.1	9.5	294	36	AV388402	AV388402 AV388402
11	41.1	9.5	667	110	BE584146	BE584146 9-8F-HA P
12	40.6	9.4	803	109	BE558685	BE558685 HV_CEB002
13	40.2	9.3	377	110	BE595391	BE595391 P11_53-F1
14	40.2	9.3	613	39	AM091035	AM091035 614070A12
15	40.0	9.3	684	107	BE387312	BE387312 601275760
16	39.8	9.2	716	110	BE636656	BE636656 rocketell
17	39.2	9.1	413	40	AM097916	AM097916 aa03e06.y
18	39.2	9.1	522	134	BE056217	BE056217 y45030E10
19	39.2	9.0	458	29	AU084061	AU084061 AU084061
20	39.0	9.0	530	29	AU084876	AU084876 AU084876
21	38.8	9.0	605	138	BE975061	BE975061 bs38b09.y
22	38.8	9.0	1011	7	AA415082	AA415082 Mg0016 RC
23	38.6	8.9	672	89	AM355915	AM355915 707016A04
24	38.4	8.9	395	140	D46616	D46616 R1C511333A
25	38.4	8.9	436	140	C97495	C97495 C97495 Rice
26	38.4	8.9	460	140	C96792	C96792 C96792 Rice
27	38.4	8.9	468	28	AJ285761	AJ285761 4A38-AAW-
28	38.4	8.9	479	88	AM244212	AM244212 687050H07
29	38.4	8.9	533	95	AM787653	AM787653 945011H09
30	38.4	8.9	550	87	AM216208	AM216208 687043C07
31	38.4	8.9	561	95	AM787652	AM787652 945011H09
32	38.4	8.9	569	23	A1649680	A1649680 486077G04
33	38.4	8.9	587	39	AM065954	AM065954 687003P09
34	38.2	8.8	458	27	A1998089	A1998089 701672121
35	38.2	8.8	498	29	AU083644	AU083644 AU083644
36	38.2	8.8	703	173	AQ096109	AQ096109 LERFJ12TR
37	38.2	8.8	360	29	AU077535	AU077535 AU077535
38	38.2	8.8	486	38	AV643609	AV643609 AV643609
39	38.2	8.8	510	38	AV645301	AV645301 AV645301
40	38.2	8.8	525	38	AV643920	AV643920 AV643920
41	38.2	8.8	815	110	BE642085	BE642085 CR12_5-B1
42	37.8	8.7	429	139	C28526	C28526 C28526 Rice
43	37.8	8.7	520	109	BE538706	BE538706 601064693
44	37.8	8.7	630	29	AU088616	AU088616 AU088616
45	37.8	8.7	645	190	CNS01213	AL101589 Drosophila

ALIGNMENTS

RESULT 1
 AA465078 120 bp mRNA
 LOCUS EST_Lubert Mtd FOX (5% oxygen) expression library Mycobacterium
 DEFINITION tuberculosis CDNA clone Lubert similar to 14 kD antigen, mRNA
 sequence.
 ACCESSION AA465078
 VERSION AA465078.1 GI:2191218
 KEYWORDS EST.

Source	Organism	Reference Authors Title	Journal MEDLINE Comment	Features Source
	Mycobacterium tuberculosis.			
	Mycobacterium tuberculosis Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium			
	1 (bases 1 to 120)			
	Imboden, P. and Schoolnik, G. K.			
	Construction and characterization of a partial Mycobacterium tuberculosis cDNA library of genes expressed at reduced oxygen tension			
	Gene 213 (1-2), 107-117 (1988)			
	Contact: Paul Imboden			
	Institute for Medical Microbiology at the Department of Clinical Research			
	University of Berne			
	DKF, Tiefenaustr. 120, 3004 Berne, Switzerland			
	Tel: ++41 31 308 80 10			
	Fax: ++41 31 308 80 28			
	Email: plimboden@imm.unibe.ch			
	position 319-420 in S79751			
	Insert Length: 99 Std Error: 10.00			
	Plate: 1 row: D column: 12.			
	Location/Qualifiers			
	1. 120			
	/organism="Mycobacterium tuberculosis"			
	/strain="pi, ATCC 51910"			
	/db_xref="taxon:1773"			
	/clone="tubert"			
	/clone_lib="Mtb FOX (5% oxygen) expression library"			
	/dev_stage="late log phase"			
	/lab_host="Escherichia coli strain Novablue (Novagen)"			
	/note="Vector: pET24a (+) (Novagen); Site 1: EcoRI; Site 2: EcoRI; total RNA isolated from liquid cultures grown to late log phase in Middlebrook 7H9/ADC in an atmosphere of 5% oxygen, 5% CO2 and 90% Nitrogen. Random fragments were amplified from total RNA by PCR with the primer upn6 (5'-GCCGAGCGCTGCAGAAATCCNNNN-3') and cloned into EcoRI site of pET24a (+). Clones from M.tb ribosomal RNA were detected by hybridization and removed."			
	BASE COUNT			
	24 a 30 c 33 g 26 t 7 others			
	ORIGIN			
	Query Match			
	Best Local Similarity 21.9%; Score 94.4; DB 7; Length 120;			
	Matches 95; Conservative 0; Mismatches 7; Indels 0; Gaps 0;			
	301 ggtcgtcgccgtagtgctgacgagacgacalttaagccactacgacaaagggcatt 360			
	1 gttgctcgtccggtatgtgctgacgagacgacatttaangccnctacnacmagggcatt 60			
	07 361 ctactgtgtcggtgtggtggtttcggaaaggaaagccacccga 402			
	DB 61 ctTACTGTGTGCGTGGCGTTTCGGAAGGAGGACCAACCNAA 102			
	RESULT 2			
	LOCUS BE214663 632 bp mRNA EST 25-JUL-2000			
	DEFINITION HV.CEBD004C06f Hordeum vulgare seedling green leaf EST library			
	HVCDNA0005 (bryslph infected & control) Hordeum vulgare cDNA clone			
	HV.CEBD004C06f, mRNA sequence.			
	ACCESSION BE214663			
	VERSION BE214663.1 GI:8902275			
	KEYWORDS EST.			
	SOURCE Hordeum vulgare			
	ORGANISM Hordeum vulgare			
	Barley.			
	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;			
	Magnoliophyta; Liliopsida; Poales; Poaceae; Hordeum.			
	1 (bases 1 to 632)			
	Wing, R., Close, T. J., Kleinbols, A., Wise, R., Begum, D., Fritsch, D., Yu			
	, Y., Anderson, H., Dale, J., Henry, D., Kernodle, S., Palmer, M., Rambo			

[illegible]


```

/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone_lib="Light Grown 1 (LG1)"
/note="Organ: 10- to 14-day-old light-grown (greenhouse)
seedlings; Vector: Lambda Zap; Site_1: XhoI; Site_2: EcoRI"
; The library was made from poly-A RNA in the cloning
vector lambda Zap II. Clones to be sequenced were
prepared by mass excision."
BASE COUNT
ORIGIN
42 a 118 c 136 g 48 t

```

Query Match	9.7%	Score 41.8	DB 88	Length 344
Best Local Similarity	53.3%	Pred. No. 0.38		
Matches 88	Conservative 0	Mismatches 77	Indels 0	Gaps 0

Qy 13 cccgltcagcgacaccgcggtccctctcccagatttctgagctyltcgcgccttc 72
||| ||| | ||| ||||| | ||| |||||
Db 104 CCCGAGCAGCAGGAGTGGAGACTTCGCCCGAGTGATGGCGCCCTGTCCGCGCTCC 163

Db 164 CCGGCTTCGGACCGCTGCTCCGACAGTACGCGCGCTCGGCTGCTGACGCGAG 222

Qy	133	atgaagaagggcgctacgaagtacgcgcgagctccggggtc	177
Db	224	gaggaacttgacgacgacgccgaggtcagacggcttccctcccgacgac	268

RESULT	8
BE593608	
LOCUS	389 bp mRNA
DEFINITION	EST 18-AUG-2000
	WS1_98_G02_b1_A002 Water-stressed 1 (WS1) Sorghum bicolor cDNA, partial sequence

ACCESSION	BES93608
VERSION	BES93608.1
KEYWORDS	GI:9848583
SOURCE	EST.
ORGANISM	sorghum, Sorghum bicolor

TITLE	JOURNAL	COMMENT
AN EST database from Sorghum: water-stressed plants unpublished (2000)		
Contact: Cordomier-Pratt MM		
AUTHORS		
Cordomier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt, L.H.		
REFERENCE		
1 (bases 1 to 369)		
Genbank accession numbers: U00001, U00002, U00003, U00004, U00005, U00006, U00007, U00008, U00009, U00010, U00011, U00012, U00013, U00014, U00015, U00016, U00017, U00018, U00019, U00020, U00021, U00022, U00023, U00024, U00025, U00026, U00027, U00028, U00029, U00030, U00031, U00032, U00033, U00034, U00035, U00036, U00037, U00038, U00039, U00040, U00041, U00042, U00043, U00044, U00045, U00046, U00047, U00048, U00049, U00050, U00051, U00052, U00053, U00054, U00055, U00056, U00057, U00058, U00059, U00060, U00061, U00062, U00063, U00064, U00065, U00066, U00067, U00068, U00069, U00070, U00071, U00072, U00073, U00074, U00075, U00076, U00077, U00078, U00079, U00080, U00081, U00082, U00083, U00084, U00085, U00086, U00087, U00088, U00089, U00090, U00091, U00092, U00093, U00094, U00095, U00096, U00097, U00098, U00099, U00100, U00101, U00102, U00103, U00104, U00105, U00106, U00107, U00108, U00109, U00110, U00111, U00112, U00113, U00114, U00115, U00116, U00117, U00118, U00119, U00120, U00121, U00122, U00123, U00124, U00125, U00126, U00127, U00128, U00129, U00130, U00131, U00132, U00133, U00134, U00135, U00136, U00137, U00138, U00139, U00140, U00141, U00142, U00143, U00144, U00145, U00146, U00147, U00148, U00149, U00150, U00151, U00152, U00153, U00154, U00155, U00156, U00157, U00158, U00159, U00160, U00161, U00162, U00163, U00164, U00165, U00166, U00167, U00168, U00169, U00170, U00171, U00172, U00173, U00174, U00175, U00176, U00177, U00178, U00179, U00180, U00181, U00182, U00183, U00184, U00185, U00186, U00187, U00188, U00189, U00190, U00191, U00192, U00193, U00194, U00195, U00196, U00197, U00198, U00199, U00200, U00201, U00202, U00203, U00204, U00205, U00206, U00207, U00208, U00209, U00210, U00211, U00212, U00213, U00214, U00215, U00216, U00217, U00218, U00219, U00220, U00221, U00222, U00223, U00224, U00225, U00226, U00227, U00228, U00229, U00230, U00231, U00232, U00233, U00234, U00235, U00236, U00237, U00238, U00239, U00240, U00241, U00242, U00243, U00244, U00245, U00246, U00247, U00248, U00249, U00250, U00251, U00252, U00253, U00254, U00255, U00256, U00257, U00258, U00259, U00260, U00261, U00262, U00263, U00264, U00265, U00266, U00267, U00268, U00269, U00270, U00271, U00272, U00273, U00274, U00275, U00276, U00277, U00278, U00279, U00280, U00281, U00282, U00283, U00284, U00285, U00286, U00287, U00288, U00289, U00290, U00291, U00292, U00293, U00294, U00295, U00296, U00297, U00298, U00299, U00300, U00301, U00302, U00303, U00304, U00305, U00306, U00307, U00308, U00309, U00310, U00311, U00312, U00313, U00314, U00315, U00316, U00317, U00318, U00319, U00320, U00321, U00322, U00323, U00324, U00325, U00326, U00327, U00328, U00329, U00330, U00331, U00332, U00333, U00334, U00335, U00336, U00337, U00338, U00339, U00340, U00341, U00342, U00343, U00344, U00345, U00346, U00347, U00348, U00349, U00350, U00351, U00352, U00353, U00354, U00355, U00356, U00357, U00358, U00359, U00360, U00361, U00362, U00363, U00364, U00365, U00366, U00367, U00368, U00369, U00370, U00371, U00372, U00373, U00374, U00375, U00376, U00377, U00378, U00379, U00380, U00381, U00382, U00383, U00384, U00385, U00386, U00387, U00388, U00389, U00390, U00391, U00392, U00393, U00394, U00395, U00396, U00397, U00398, U00399, U00400, U00401, U00402, U00403, U00404, U00405, U00406, U00407, U00408, U00409, U00410, U00411, U00412, U00413, U00414, U00415, U00416, U00417, U00418, U00419, U00420, U00421, U00422, U00423, U00424, U00425, U00426, U00427, U00428, U00429, U00430, U00431, U00432, U00433, U00434, U00435, U00436, U00437, U00438, U00439, U00440, U00441, U00442, U00443, U00444, U00445, U00446, U00447, U00448, U00449, U00450, U00451, U00452, U00453, U00454, U00455, U00456, U00457, U00458, U00459, U00460, U00461, U00462, U00463, U00464, U00465, U00466, U00467, U00468, U00469, U00470, U00471, U00472, U00473, U00474, U00475, U00476, U00477, U00478, U00479, U00480, U00481, U00482, U00483, U00484, U00485, U00486, U00487, U00488, U00489, U00490, U00491, U00492, U00493, U00494, U00495, U00496, U00497, U00498, U00499, U00500, U00501, U00502, U00503, U00504, U00505, U00506, U00507, U00508, U00509, U00510, U00511, U00512, U00513, U00514, U00515, U00516, U00517, U00518, U00519, U00520, U00521, U00522, U00523, U00524, U00525, U00526, U00527, U00528, U00529, U00530, U00531, U00532, U00533, U00534, U00535, U00536, U00537, U00538, U00539, U00540, U00541, U00542, U00543, U00544, U00545, U00546, U00547, U00548, U0		

Department of Botany
The University of Georgia
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 542 1805
Email: mmpratt@uga.edu
Sequences have been trimmed to exclude POLYA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.
Seq primer: JEN REV
High quality sequence stop: 383
POLYA-NO.

FEATURES	Location/Qualifiers
source	1. .389

```

SOURCE
1:1501
/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone_1lb="Water-stressed 1 (WS1)"
/note="Organ: Mix of 5-week old plants on days 7 & 8 after
water was withheld; Vector: lambda zap; Site 1: XhoI;
Site 2: EcoRI; The library was made from poly-A RNA in the
cloning vector lambda zap II. Clones to be sequenced were
prepared by mass excision."
BASE COUNT
56 a 129 c 153 g 51 t
ORIGIN

```

[illegible]

RESULT 9
BE552631

DEFINITION	946082H01.y1 946 - tassal primordium prepared by Schmidt lab Zea
ACCESSION	mays cDNA, mRNA sequence.
BE552631	
Accession #	0701000

KEYWORDS	EST.	1
SOURCE	Zea mays.	
ORGANISM	Zea mays	

REFERENCE	AUTHORS	TITLE
1 (bases 1 to 531)	Walbot, V.	Maize ESTs from various cDNA libraries sequenced at Stanford University
2 (bases 532 to 1000)	Walbot, V.	Maize ESTs from various cDNA libraries sequenced at Stanford University

COMMENT
Contact: Waldo V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227

```

features
  source
    fax: 001 252 3221
    email: walbot@stanford.edu
    plate: 946082 row: H column: 01.
    location/Qualifiers
      1. 531
    /organism="Zea mays"
    /cultivar="OH43"
    /db_xref="taxon:4577"

```

```

/clone_id= 340 -tassel primordium prepared by Schmidt
lab=
/tissue_type="tassels"
/dev_stage="just after the transition from vegetative to
inflorescence development"
/lab_host="XLOLR"
/note="Organ: tassels; Vector: HybriZAP; Site_1: EcoRI;
Site_2: XhoI; George Chuck dissected immature tassels
between 1mm and 3mm. Sharon Stanfield prepared the cDNA
library in HybriZAP. Sample insert size range was 350 bp
to 3 kb with a 1 kb average."
218 c 171 g 70 t
72 a

```

ORIGIN

	Query Match	9.6%;	Score 41.6;	DB 109;	Length 531;	-
	Best Local Similarity	50.5%;	Pred. No. 0.44;			
	Matches 101; Conservative	0;	Mismatches 99;	Indels 0;	Gaps 0;	
QY	55 gacgtgctcgagcgccttcgccgatcatctggccggaactcggcccaccttcgaaccggltg	114				
D8	332 GCGAGGTCGAGGGACTGCGCCCACTCCGCCGGGGGTCCGGTGAATCCTTGACCTTTTTC	391				
OY	115 atggagatcgaaagcaagaalaaagaaggqgcgtcacagqtgtaacgcgcgaagtccccgg	174				


```

OY 175 gtccaccccaagaagcgtccacattatgctcgcgatgtcagctaccatcaagcc 234
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 452 cttaattaccttcattgctgacgtgacgtgacgacgacgcttttgagacgattgacgac 511
OY 235 gagcgaccgagcagaaga 254
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 512 GACGACGAGGAGGAGGACGA 531

RESULT 10
AV388402 294 bp mRNA EST 29-SEP-2000
LOCUS AV388402
DEFINITION AV388402 Chlamydomonas reinhardtii C9 Chlamydomonas reinhardtii
ACCESSION CDNA clone CM033D05_r, mRNA sequence.
VERSION AV388402.1 GI:6542618
KEYWORDS EST.
SOURCE Chlamydomonas reinhardtii.
ORGANISM Chlamydomonas reinhardtii.
REFERENCE Chlamydomonadaceae; Chlamydomonas.
AUTHORS Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
TITLE Chlamydomonadaceae; Chlamydomonas.
1 (bases 1 to 294)
Asamizu, E., Nakamura, Y., Sato, S., Fukuzawa, H. and Tabata, S.
A large scale structural analysis of cDNAs in a unicellular green
algae, Chlamydomonas reinhardtii. I. Generation of 3433
non-redundant expressed sequence tags
JOURNAL DNA Res. 6 (6), 369-373 (1999)
MEDLINE 20152988
COMMENT Contact: Yasukazu Nakamura
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: ynakam@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.
FEATURES
    source
        1..294
            /organism="Chlamydomonas reinhardtii"
            /strain="C9"
            /db_xref="taxon:3055"
            /clone="CM033D05_r"
            /clone_lib="Chlamydomonas reinhardtii C9"
            /dev_stage="photoautotrophic growth"
            /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
            XhoI"
BASE COUNT 54 a 86 c 118 g 34 t 2 others
ORIGIN
Query Match 9.5%; Score 41; DB 36; Length 294;
Best Local Similarity 48.5%; Pred. No. 0.61;
Matches 110; Conservative 0; Mismatches 117; Indels 0; Gaps 0;
OY 102 cgacacccggttgaatgcggtgaaagacagatgaagaaggcgctacgaagctacgcgc 161
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 67 CGACCCGCGCTGTGATGAGCTGTGACAGAGATGGCGGATCGCCCGACGACGCCGCCCTC 126
OY 162 ggaagctcccgagggtgcagaccgcagacaagagctccacattatgctcgcgattgaact 221
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 127 GCTGTCGGGGGGGAGGAGCGACGACGACGACGACGACGCGCTGCGCTGAGCTGGAGGC 186
OY 222 gaccatcaagcgagcgacgcagcagacagagacttagacgctgctggaattgcgta 281
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 187 GCGCGTGGCATGAGACTCTCAGAGACCAAGAGCGCGCCGCAAGCTCCAAAGGCTTCAA 246
OY 282 cggctccctcgttcgacaggtgctgcctccggttaggtgtgtgaagcga 328
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 247 GCAGGCTGAGGTCATGATGTGTGTGAGGCGCGGATGNNMGAGG 293

RESULT 11
BE584146 667 bp mRNA EST 16-AUG-2000
LOCUS BE584146
DEFINITION 9-8F-HA Psoljaeha Glycine max/Phytophthora sojae mixed EST library

```

```

ACCESSION BE584146
VERSION BE584146.1 GI:9835095
KEYWORDS Glycine max/Phytophthora sojae mixed EST library.
SOURCE Glycine max/Phytophthora sojae mixed EST library.
ORGANISM Eukaryota; mixed EST libraries.
REFERENCE 1 (bases 1 to 667)
AUTHORS Outub,D., Hrader,P.T., Sobral,B.W.S. and Gijzen,M.
TITLE Comparative analysis of expressed sequences in Phytophthora sojae
JOURNAL Plant Physiol. 123 (1), 243-254 (2000)
MEDLINE 20267956
COMMENT Contact: Gijzen M
Agriculture and Agri-Food Canada
1391 Sandford Street, London, Ontario, Canada N5V 4T3
Tel: 519 457 1470
Fax: 519 457 3997
Email: gijzenm@em.agr.ca
Transcript obtained from mixed plant-pathogen interaction culture.
FEATURES
    source
        1..667
            /organism="Glycine max/Phytophthora sojae mixed EST
            library"
            /strain="Phytophthora sojae race 2 strain p6497"
            /cultivar="Glycine max cultivar Harosoy"
            /db_xref="taxon:135715"
            /clone_lib="Psoljaeha"
            /tissue_type="Plant hypocotyls infected with Phytophthora
            sojae"
            /dev_stage="8 d old etiolated hypocotyls 48 h
            post-inoculation with Phytophthora sojae zoospores"
            /lab_host="E. coli strain XL0LR"
            /note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI; This
            cDNA library was constructed from polyA+ enriched mRNA
            from etiolated hypocotyls 48 h post-inoculation with
            Phytophthora sojae zoospores. Complementary DNA was
            synthesized from mRNA using an XhoI-poly(dT)
            linker-primer. EcoRI adapters were ligated to the
            blunt-ended cDNA fragments and the products were digested
            with XhoI for directional cloning into lambda ZAP Express
            vector. This lambda library was amplified once using E.
            coli host strain XL1 Blue MRF'. Inserts were then
            subcloned by mass excision using Exsist helper phage
            for conversion into phagemid vector pBK-CMV in E. coli
            host strain XL0LR. Sequenced using T3 primer: 5' ATT AAC
            CCT CAC TAA AGG GA 3'."
BASE COUNT 153 a 176 c 204 g 133 t 1 others
ORIGIN
Query Match 9.5%; Score 41; DB 110; Length 667;
Best Local Similarity 45.5%; Pred. No. 0.64;
Matches 146; Conservative 0; Mismatches 175; Indels 0; Gaps 0;
OY 101 tgaacacccggttgaatgcggtgaaagacagatgaagaaggcgctcagagtgacgc 160
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 29 TCGCACACCGCTGCTGCTGTGGGGTACTCTGAAGCAAGCTCCACCCGACTTCCGGT 88
OY 161 cggagctcccgagggtgcagaccgcagacaagacgtcacattatgctcgcgattgcaac 220
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 89 AGCGCTTCCCAAGACGGCCACACAGAAAGAGGCAAGAGGCGCTGCTCTGCTGAGC 148
OY 221 tgaacatcaagcgagcgacgcagcagaaggaacttagacgctcgcgtaattgcgct 280
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 149 CGGCCAAGAGCCACGACGACGAGGTACTGTGACGACGAGAGTTCGCCGAGTGGACAC 208
OY 281 acggttcctcgttcgacaggtgctgcgctgaggtgtgtcgaagagacatattag 340
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 209 AGAAGGTGCTGTTCATGAGCGGTGCTGTCGATGATGTCGACCAAGCAGTTCGCTG 268
OY 341 ccacatcaagaaggcattcttactgtgtcgttgagcggttcggaagggaagcacaagc 400
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 269 ACTTCAATGAGACGCTCAAGGCCGAGATCTCGAAGAGAGACTTCGCTAAGAAAGCTGACCA 328

```


Db 124 CCGGGCTTCCCGGACCGTGGCTCCGAGATGACGGCGCTCGGGCTGCTCGACGCCGAG 183

Qy 133 a tgaagagagggcgtacagatgacgcgcgcagcattcccgagtc 177

Db 184 GAGGACCTGGACGACGCCGAGGTCGACGGCTTCCTCCCGACGAC 228

RESULT 14

LOCUS	AM091035	613 bp	mRNA	EST	18-OCT-1999	
DEFINITION	614070A12.Y1 614 - root cDNA library from Malbot Lab Zea mays cDNA.					
ACCESSION	AM091035	mRNA sequence.				
VERSION	AM091035.1	GI:6056630				
KEYWORDS	EST.					
SOURCE	Zea mays.					

FEATURES

Maize ESTs from various cDNA libraries sequenced at Stanford University
Unpublished (1999)
Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 614070 row: A column: 12.

QY	235	gagcgacacgagacgaaggaacttaacgctgcgttcggaattcgcgtacaggttcttcgtt	294
Db	506	GAGCAGGAGCAGAGAGACGGACACCTGCTCCGCGCTGGAGCGAGCAGCGGAGAGTTCCTG	447
QY	295	cgcacggtgtctgcgtccgcgtagtgcctgcagagagacataagtcgacctacgacaag	354
Db	446	CGCAGGTTCCGGGCTCCCGAGAACGCCAACAAGGAGCAGATCAACGCCGCGCATGGAAAC	387
QY	355	ggaattactatgtctgcgtgttcgagtcgttcgcgaagggagagccaacccgaaagca	407
Db	386	GGCTGCTTACATGCATCTGTGCGCAGAGAGAGACCCCAAGAGACCTTGAAGTGA	334

RESULT 15

LOCUS	BE387312	684 bp	mRNA	21-JUL-2000
DEFINITION	6012257670F1 NIH_MGC_20		Homo sapiens cDNA clone IMAGE:3616574	5'
ACCESSION	mRNA sequence.			
VERSION	BE387312			
CLONE	BE387312.1	GI:9332677		

KEYWORDS
SOURCE
ORGANI

EST.
human.

SOURCE

ORGANISMS

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 684)

REFERENCE

AUTHORS
 TITLE
 JOURNAL
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 http://www.ncbi.nlm.nih.gov/mgc/

COMMENT

Tel: (301) 496-1550
 Email: Robert_Strausberg@nih.gov
 Tissue Procurement: ATCC/DC/DTF
 cDNA Library Preparation: Ling Hong/Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Invitae Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium at: image.llnl.gov
 Plate: L12CM282 row: h column: 15
 High quality sequence start: 71
 High quality sequence stop: 661.

FEATURES

BASE COUNT
ORIGIN

170 a	205 c	186 g	122 t	1 others
-------	-------	-------	-------	----------

QY	88	ctccgccaccactccgcacaccggttatgagcgttggaagacagatgaaagagggcgc	147
Db	297	CTGAGAGACGCTCAACGCACATGTACCTCTGCTGGAGGAGACGGAGAAACAG66CCGTCCG	356
QY	148	tacgaggtacgcgcgcagctcccgaggctcgaccccgcaagaagacgtccacatattgctc	207
Db	357	CGGGCGCGGATCGAGAGAGGGGGCGGCTTTCGACACTTCATCACACTTCCGCAACCCGTGG	416
QY	208	cgcgactgtctcagctgacacatcaagcccgagcgcacgcgagagaaaggacttagagcgtcgc	267
Db	417	GTGATGTGGAGAGCTGACCATGCTGGGAGAGATCACCCACCTGCAGGGGCATCATGACGAT	476
QY	268	tcggg	271
Db	477	TTTGG	480

Search completed: January 12, 2001, 18:01:41
Job time: 21237 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 12, 2001, 20:14:13 ; Search time 153.29 Seconds
(without alignments)
454,180 Million cell updates/sec

Title: US-09-461-774-7

Perfect score: 432
Sequence: 1 gccacacccctcccgctca.....agatcgcgtccaccactga 432

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 280836 seqs, 80580151 residues

tal number of hits satisfying chosen parameters: 561672

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

Issued_Patents_NA:*
1: /cgn2_6/pdata/1/lna/5A.COMB.seq:*
2: /cgn2_6/pdata/1/lna/5B.COMB.seq:*
3: /cgn2_6/pdata/1/lna/6.COMB.seq:*
4: /cgn2_6/pdata/1/lna/PCRTUS.COMB.seq:*
5: /cgn2_6/pdata/1/lna/Backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	39.4	9.1	972	4	PCT-US95-04801-2
2	39	9.0	4897	5	5196516-7
3	38.2	8.8	1108	4	PCT-US93-03035-1
4	36.4	8.4	3231	1	US-08-074-121-4
5	36.4	8.4	3231	4	PCT-US94-06447-4
6	36.2	8.4	1482	3	US-08-660-645A-6
7	36.2	8.4	1482	3	US-09-298-718-6
8	36	8.3	1800	1	US-08-260-202A-10
9	36	8.3	1800	1	US-08-017-114-10
10	36	8.3	1800	3	US-08-505-307-10
11	36	8.3	1800	4	PCT-US94-02034-10
12	35.2	8.1	2634	1	US-08-196-218-31
13	35.2	8.1	2634	1	US-08-681-953-31
14	34.4	8.0	1785	2	US-08-910-856-9
15	34.4	8.0	3187	2	US-08-910-856-9
16	34.4	8.0	3187	2	US-08-910-856-10
17	34.2	7.9	1138	3	US-08-581-148C-3
18	33.8	7.8	620	2	US-08-757-653-135
19	33.8	7.8	620	2	US-08-757-653-136
20	33.8	7.8	620	2	US-08-757-653-137
21	33.8	7.8	620	2	US-08-757-653-138
22	33.8	7.8	620	2	US-08-757-653-139
23	33.8	7.8	620	2	US-08-757-653-140
24	33.8	7.8	917	2	US-08-951-822-1
25	33.8	7.8	970	1	US-08-250-030-1
26	33.8	7.8	970	4	PCT-US95-06790-1
27	33.8	7.8	44377	2	US-08-804-227C-7
28	33.8	7.8	44377	2	US-08-804-198-1

29	33.6	7.8	3744	2	US-08-348-353-16	Sequence 16, Appl
30	33.6	7.8	3744	2	US-08-465-965-16	Sequence 16, Appl
31	33.6	7.8	3744	3	US-08-465-966-16	Sequence 16, Appl
32	33.4	7.7	1450	1	US-07-923-692C-5	Sequence 5, Appl
33	33.4	7.7	1450	1	US-08-184-237-5	Sequence 5, Appl
34	33.4	7.7	1450	2	US-08-482-920-5	Sequence 5, Appl
35	33.4	7.7	1450	3	US-08-484-341-5	Sequence 5, Appl
36	33.4	7.7	20235	1	US-07-642-736C-3	Sequence 3, Appl
37	33.4	7.7	20235	3	US-08-439-009A-3	Sequence 3, Appl
38	33.2	7.7	1052	2	US-08-403-852D-10	Sequence 10, Appl
39	33.2	7.7	1052	3	US-08-510-646B-10	Sequence 10, Appl
40	33.2	7.7	38506	3	US-09-320-878-19	Sequence 19, Appl
41	33	7.6	285	5	5244792-8	Patent No. 5244792
42	33	7.6	1334	2	US-08-916-901-2	Sequence 2, Appl
43	33	7.6	30001	1	US-08-125-468-1	Sequence 1, Appl
44	33	7.6	30001	2	US-08-474-933-1	Sequence 1, Appl
45	32.8	7.6	43280	2	US-08-804-227C-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
PCT-US95-04801-2
Sequence 2, Application PC/TUS9504801
GENERAL INFORMATION:
APPLICANT: Martin, Juan F.
APPLICANT: Coque, Juan R.
APPLICANT: Enguita, Francisco J.
APPLICANT: Fuente, Juan L.
APPLICANT: Larena, Francisco J.
APPLICANT: Liras, Paloma
TITLE OF INVENTION: DNA ENCODING CEPHARYCIN BIOSYNTHESIS
TITLE OF INVENTION: LATE GENES
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: John W. Wallen III
STREET: P.O. Box 2000
CITY: Rahway
STATE: New Jersey
COUNTRY: USA
ZIP: 07065
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/04801
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Wallen III, John W.
REGISTRATION NUMBER: 35,403
REFERENCE/DOCKET NUMBER: 19179
TELECOMMUNICATION INFORMATION:
TELEPHONE: (908) 594-3905
TELEFAX: (908) 594-4720
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 972 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
PCT-US95-04801-2

Query Match 9.1%; Score 39.4; DB 4; Length 972;
Best Local Similarity 47.1%; Pred. No. 0.06;
Matches 121; Conservative 0; Mismatches 136; Indels 0; Gaps 0;

QY 120 gctggaagcagatgaagaaggggcgcacagagacgcggagctcccggggacgca 179

```
Db 45 GCGCAGAGAGAGTTCCGCGAGTGGCCCGCGGGGTCTTCTACCTCACCGGGTACGG 104
Qy 180 ccccgagaagagcgtccacattatgctcgcgagatgctcagcttgacatcaagagcgagcg 239
Db 105 GCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 164
Qy 240 caccgagcagaagagacttagacggtcgtcgaaatcgcgtacggttcctcgtcgtcgcac 299
Db 165 CACGCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 224
Qy 300 ggtgctgctcgcggtgagtgctgctgacgaggaagacattaaagccacctaagagagcgat 359
Db 225 GCGCCTGAGAGCGGAAAGACCGCCGAGTCCACCAACACCGCAGCTACACTCTC 284
Qy 360 tctactcgtcgttgg 376
Db 285 CATGCTGCTACTGATGG 301
```

```
ULT 2
5196516-7
; Patent No. 5196516
; APPLICANT: SCHREURS, CHRISTA S.; METTENLEITER, THOMAS C.
; SIMON, ARTUR J.; LUKAS, NOEMI; RZIH, HANNS J.
; TITLE OF INVENTION: PSEUDORABIES VIRUS VACCINE
; NUMBER OF SEQUENCES: 8
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/383,833
; FILING DATE: 21-JUL-1989
; SEQ ID NO: 7
; LENGTH: 4897
```

```
Query Match 9.08; Score 39; DB 5; Length 4897;
Best Local Similarity 48.8%; Pred. No. 0.11;
Matches 105; Conservative 0; Mismatches 110; Indels 0; Gaps 0;
```

```
Qy 41 tcccgaggtttctgagcttctcgcgcttcgcgcttcgctcgcgagcttcgcgagcgcac 100
Db 385 tctggtggtcgtcgcgcttcgctcgcgcttcgctcgcgcttcgctcgcgagcgcac 3344
Qy 101 tgcagaccggttgatcgcgtcgaagagagagagagagagagagagagagagagagagag 160
Db 3445 gtccgaccccatgaaagagagagagagagagagagagagagagagagagagagagagag 3404
Qy 161 cgaagcttcgcggtcgcagcccgacagagagagagagagagagagagagagagagagag 220
Db 3405 tgcagagagagagagagagagagagagagagagagagagagagagagagagagagagag 3464
Qy 221 tgaccatcaagagcgcgagcgcagcgcagcgcagcgcagcgcagcgcagcgcagcgcag 255
Db 3465 tgtccatcgtcgtcgcgcttcgagcagcagcagcagcagcagcagcagcagcagcagcag 3499
```

```
RESULT 3
PCT-US93-03035-1
; Sequence 1, Application PC/TUS9303035
; GENERAL INFORMATION:
; APPLICANT: ABBOTT LABORATORIES
; TITLE OF INVENTION: PURIFIED THERMOSTABLE ENDONUCLEASE
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ABBOTT LABORATORIES
; STREET: D-377 APD, ONE ABBOTT PARK ROAD
; CITY: ABBOTT PARK
; STATE: ILLINOIS
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
```

```
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/03035
; FILING DATE: 19930330
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/869,306
; FILING DATE: 16-APR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/860,702
; FILING DATE: 31-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: BRAINARD, THOMAS D
; REGISTRATION NUMBER: 32,459
; REFERENCE/DOCKET NUMBER: 5145.PC.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 708-937-4884
; TELEFAX: 708-937-2623
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1108 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 114..926
; PCT-US93-03035-1
```

```
Query Match 8.8%; Score 38.2; DB 4; Length 1108;
Best Local Similarity 48.0%; Pred. No. 0.13;
Matches 109; Conservative 0; Mismatches 118; Indels 0; Gaps 0;
```

```
Qy 23 gcaacccgagtcctcctcctccgagcttctgagctcgtcgcgcttcgcgcttcgcgcttcgcg 82
Db 235 GCGCGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 294
Qy 83 ccgagctcgcgagccacacttcgacacccggttgatgctgctggaagagagagagagagag 142
Db 295 CCGGGGCGCTCCCGCGGTATCCACGCTTACTGATCTGATCTGATCTGATCTGATCTGATCTG 154
Qy 143 ggcgtacgagagagagagagagagagagagagagagagagagagagagagagagagagag 202
Db 355 AGCTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 414
Qy 203 tgtccgagatgctcagctgacacatcaagagcgcagcgcagcgcagcgcagcgcagcgcag 249
Db 415 GCGTGGAGTACGTGCTGCTCACCCTCGGCTCGGCGCGCGCGCGCGCGCGCGCGCGCGCG 461
```

```
RESULT 4
US-08-074-121-4
; Sequence 4, Application US/08074121
; Patent No. 5767362
; GENERAL INFORMATION:
; APPLICANT: Best, Elaine
; APPLICANT: Knaut, Vito C.
; TITLE OF INVENTION: Methods and Compositions for Modulating
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 2200 Sand Hill Road, Suite 100
; CITY: Menlo Park
; STATE: California
; COUNTRY: US
; ZIP: 94025
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/074.121
FILING DATE: 08-JUN-1993
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Rae-Ventler, Barbara
REGISTRATION NUMBER: 32,750
REFERENCE/DOCKET NUMBER: 05938/043001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 854-5277
TELEFAX: (415) 854-0875
TELEX: 200154
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 3231 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 861..1328
FEATURE:
NAME/KEY: CDS
LOCATION: 1349..2695
US-08-074-121-4

Query Match 8.4%; Score 36.4; DB 1; Length 3231;
Best Local Similarity 50.9%; Pred. No. 0.49;
Matches 113; Conservative 0; Mismatches 106; Indels 3; Gaps 1;

QY 1 gccacacccctccggttaagcgcaaccgcggtccctctcccgagtttctgagctg 60
DB 1826 gccgcccggggggggggtggcgccgacatgctggtttacgacgagtcgacgac 1885
QY 61 ttccgagccctccgctcattcgccgagctcggccaccttcgacaccccggttgatgcg 120
DB 1886 aagtgccgacactgacccgacacgagcgccggcgcttcggcgaacccgattggtctac 1945
QY 121 ctggaagacgagatgaagagggcgctacgagtgacgagcgagcttcccgggatcgac 180
DB 1946 ctggagaaattcttgacacacccggcgccgagtggaagttcagagtgcttccga---cgcc 2002
181 ccgacaagagcgtccacattatgttcgcgagatgttcagctg 222
2003 cagggcgaacgcatccactcgcgcgacgacgactgctccctg 2044

RESULT 5
PCT-US94-06447-4
Sequence 4, Application PC/TUS9406447
GENERAL INFORMATION:
APPLICANT: Calgene, Inc.
TITLE OF INVENTION: Methods and Compositions for Modulating
TITLE OF INVENTION: Lipid Content of Plant Tissues
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Well, Gotshal & Manges
STREET: 2882 Sand Hill Road, Suite 280
CITY: Menlo Park
STATE: California
COUNTRY: US
ZIP: 94025
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US94/06447
FILING DATE: 06-JUN-1994
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Rae-Ventler, Barbara
REGISTRATION NUMBER: 32,750
REFERENCE/DOCKET NUMBER: CGNE-097/WO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 926-6200
TELEFAX: (415) 854-3713
TELEX:
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 3231 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 861..1328
FEATURE:
NAME/KEY: CDS
LOCATION: 1349..2695
PCT-US94-06447-4

Query Match 8.4%; Score 36.4; DB 4; Length 3231;
Best Local Similarity 50.9%; Pred. No. 0.49;
Matches 113; Conservative 0; Mismatches 106; Indels 3; Gaps 1;

QY 1 gccacacccctccggttaagcgcaaccgcggtccctctcccgagtttctgagctg 60
DB 1826 gccgcccggggggggggtggcgccgacatgctggtttacgacgagtcgacgac 1885
QY 61 ttccgagccctccgctcattcgccgagctcggccaccttcgacaccccggttgatgcg 120
DB 1886 aagtgccgacactgacccgacacgagcgccggcgcttcggcgaacccgattggtctac 1945
QY 121 ctggaagacgagatgaagagggcgctacgagtgacgagcgagagcttcccgggatcgac 180
DB 1946 ctggagaaattcttgacacacccggcgccgagtggaagttcagagtgcttccga---cgcc 2002
QY 181 ccgacaagagcgtccacattatgttcgcgagatgttcagctg 222
DB 2003 cagggcgaacgcatccactcgcgcgacgacgactgctccctg 2044

RESULT 6
US-08-660-645A-6
Sequence 6, Application US/08660645A
Patent No. 6087152
GENERAL INFORMATION:
APPLICANT: Hohmann, Hans-Peter
APPLICANT: Pasamontes, Luis
APPLICANT: Tessier, Michel
APPLICANT: van Loon, Adolphus
TITLE OF INVENTION: FERMENTATIVE CAROTENOID PRODUCTION
NUMBER OF SEQUENCES: 47
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hoffmann-La Roche Inc.
STREET: 340 Kingsland Street
CITY: Nutley
STATE: NJ
COUNTRY: USA
ZIP: 07110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/660,645A

FILING DATE: 07-JUN-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 95108888.9
FILING DATE: 09-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Pokras, Bruce A.
REGISTRATION NUMBER: 32,748
REFERENCE/DOCKET NUMBER: RAN 6002/170
TELECOMMUNICATION INFORMATION:
TELEPHONE: (201) 235-5801
TELEFAX: (201) 235-2363
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1482 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
08-660-645A-6

Query Match 8.4%; Score 36.2; DB 3; Length 1482;
Best Local Similarity 53.1%; Pred. No. 0.46; Mismatches 68; Indels 0; Gaps 0;
Matches 77; Conservative 0;

QY 1 gccacacccctccggttcagcgccacccgagtcctctcccccaggtttctgagctg 60
DB 880 GCCGCGATCCTGAACCGCGACGCGTGTGATGCTGCTGCTGCTGCTGCTGCTG 939
QY 61 ttccgagccttcctccgtatcgcgcgagctccgcgcaccccttcgacacccggttgatcg 120
DB 940 TCCAAGGCGCCCGAAGACCTGCGCCACACAGCGTATCTTGGCGCCGCGCTACAAGGG 999
QY 121 ctggaagcagagatgaagaggagc 145
DB 1000 CTGCTGACGAGATCTTCAACGGGC 1024

RESULT 7

US-09-298-718-6
Sequence 6, Application US/09298718
Patent No. 6124113
GENERAL INFORMATION:
APPLICANT: Hohmann, Hans-Peter
APPLICANT: Pasamontes, Luis
APPLICANT: Tessier, Michel
APPLICANT: van Loon, Adolpinus
TITLE OF INVENTION: FERMENTATIVE CAROTENOID PRODUCTION
NUMBER OF SEQUENCES: 47
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hoffmann-La Roche Inc.
STREET: 340 Kingsland Street
CITY: Nutley
STATE: NJ
COUNTRY: USA
ZIP: 07110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/298,718
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/660,645
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Pokras, Bruce A.
REGISTRATION NUMBER: 32,748
REFERENCE/DOCKET NUMBER: RAN 6002/170

TELECOMMUNICATION INFORMATION:
TELEPHONE: (201) 235-5801
TELEFAX: (201) 235-2363
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1482 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-09-298-718-6

Query Match 8.4%; Score 36.2; DB 3; Length 1482;
Best Local Similarity 53.1%; Pred. No. 0.46; Mismatches 68; Indels 0; Gaps 0;
Matches 77; Conservative 0;

QY 1 gccacacccctccggttcagcgccacccgagtcctctcccccaggtttctgagctg 60
DB 880 GCCGCGATCCTGAACCGCGACGCGTGTGATGCTGCTGCTGCTGCTGCTGCTG 939
QY 61 ttccgagccttcctccgtatcgcgcgagctccgcgcaccccttcgacacccggttgatcg 120
DB 940 TCCAAGGCGCCCGAAGACCTGCGCCACACAGCGTATCTTGGCGCCGCGCTACAAGGG 999
QY 121 ctggaagcagagatgaagaggagc 145
DB 1000 CTGCTGACGAGATCTTCAACGGGC 1024

RESULT 8

US-08-260-202A-10
Sequence 10, Application US/08260202A
Patent No. 5573910
GENERAL INFORMATION:
APPLICANT: Deretic, Vojo
APPLICANT: Martin, Daniel W.
TITLE OF INVENTION: DETECTION OF CONVERSION TO MUCOIDY IN
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold, White & Durkee
STREET: P. O. Box 4433
CITY: Houston
STATE: TX
COUNTRY: USA
ZIP: 77210-4433
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/260,202A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/260,202
FILING DATE: 15-JUN-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/017,114
FILING DATE: 12-FEB-1993
ATTORNEY/AGENT INFORMATION:
NAME: Hodgins, Daniel S.
REGISTRATION NUMBER: 31,026
REFERENCE/DOCKET NUMBER: UTSK:221\HOD
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 418-3000
TELEFAX: (512) 474-7577
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 1800 base pairs
TYPE: nucleic acid


```

1      RESULT 13
2      US-08-681-953-31
3      ; Sequence 31, Application US/08681953
4      ; Patent No. 5710032
5      ; GENERAL INFORMATION:
6      ; APPLICANT: Piepersberg, Wolfgang
7      ; APPLICANT: Stockmann, Michael
8      ; APPLICANT: Taleghani, Kamriz Mansouri
9      ; APPLICANT: Distler, Susanne
10     ; APPLICANT: Grabley, Susanne
11     ; APPLICANT: Sichel, Petra
12     ; APPLICANT: Brau, Barbara
13     ; TITLE OF INVENTION: Secondary-Metabolite Biosynthesis Genes
14     ; TITLE OF INVENTION: From Actinomycetes, Method of Isolating Them, and Their
15     ; TITLE OF INVENTION: Use.
16     ; NUMBER OF SEQUENCES: 34
17     ; CORRESPONDENCE ADDRESSES:
18     ; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
19     ; ADDRESSEE: Dunner
20     ; STREET: 1300 I Street, N.W.
21     ; CITY: Washington
22     ; STATE: D.C.
23     ; COUNTRY: United States
24     ; ZIP: 20005-3315
25     ; COMPUTER READABLE FORM:
26     ; MEDIUM TYPE: Floppy disk
27     ; COMPUTER: IBM PC compatible
28     ; OPERATING SYSTEM: PC-DOS/MS-DOS
29     ; SOFTWARE: PatentIn Release #1.0, Version #1.25
30     ; CURRENT APPLICATION DATA:
31     ; APPLICATION NUMBER: US/08/681,953
32     ; FILING DATE: 30-JUL-1996
33     ; CLASSIFICATION: 435
34     ; PRIOR APPLICATION DATA:
35     ; APPLICATION NUMBER: US 08/196,218
36     ; FILING DATE: 25-AUG-1994
37     ; ATTORNEY/AGENT INFORMATION:
38     ; NAME: Ogden, Stasia L.
39     ; REGISTRATION NUMBER: 36,228
40     ; REFERENCE/DOCKET NUMBER: 02481.1372-00000
41     ; TELECOMMUNICATION INFORMATION:
42     ; TELEPHONE: 202-408-4000
43     ; TELEFAX: 202-408-4400
44     ; INFORMATION FOR SEQ ID NO: 31:
45     ; SEQUENCE CHARACTERISTICS:
46     ; LENGTH: 2634 base pairs
47     ; TYPE: nucleic acid
48     ; STRANDEDNESS: single
49     ; TOPOLOGY: linear
50     ; MOLECULE TYPE: DNA (genomic)
51     ; FEATURE:
52     ; NAME/KEY: CDS
53     ; LOCATION: 3..401
54     ; FEATURE:
55     ; NAME/KEY: CDS
56     ; LOCATION: 416..1531
57     ; FEATURE:
58     ; NAME/KEY: CDS
59     ; LOCATION: 1561..2625
60     ; US-08-681-953-31

```

Query Match	8.1%	Score 35.2	DB 1	Length 2634
Best Local Similarity	49.0%	Pred No. 0.95		
Matches 94	Conservative 0	Mismatches 98	Indels 0	Gaps 0
QY	82	gcgcgaactccgcgcacacttcgcacaccgcggttgatgctgcgtgaagaacagatgtaagaag	141	
DB	1022	ggcgcagacggcgcgcgcctcttgcgtgacgtrccggagccggcggtgcgcccaaggacaccttgcgcac	1081	
QY	142	ggcgcctacagagtaacgcgcgcgagcttccgcggggttcgacccgcacaagaagcttccacatt	201	

Db 1082 AAGAACTCGAATTTCGACGCGGCTCGCCAAAGACGTCGCGCCCTCGGAACTGACGTCG 1141

Qy 202 atggttcgogagaatgctgcgtgcacacatcaagcgcagcgcacccgagagaaagacttagac 261

Db 1142 ATCTCTGCCCGCGCGGAGAGCGTGGCCGACGCCCTGACAGACGAACCTGCCGGGATCCGG 1201

Qy 262 ggtcgcctcgaa 273

Db 1202 GCCGCGCTGGGCA 1213

```

1      RESULT 14
2      US-08-910-856-1
3      : Sequence 1, Application US/08910856
4      : Patent No. 5981257
5      :
6      : GENERAL INFORMATION:
7      : APPLICANT: FUKUI, TOSHIKAI
8      : APPLICANT: DOI, YOSHIHARU
9      : TITLE OF INVENTION: POLYESTER SYNTHASE GENE AND PROCESS
10     : TITLE OF INVENTION: FOR PRODUCING POLYESTER
11     : NUMBER OF SEQUENCES: 16
12     :
13     : CORRESPONDENCE ADDRESS:
14     : ADDRESSEE: FISH & RICHARDSON P.C.
15     : STREET: 4225 EXECUTIVE SQUARE, SUITE 1400
16     : CITY: LA JOLLA
17     : STATE: CA
18     : COUNTRY: USA
19     : ZIP: 92037
20     :
21     : COMPUTER READABLE FORM:
22     : MEDIUM TYPE: Floppy disk
23     : COMPUTER: IBM PC compatible
24     : OPERATING SYSTEM: PC-DOS/MS-DOS
25     : SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
26     :
27     : CURRENT APPLICATION DATA:
28     : APPLICATION NUMBER: US/08/910, 856
29     : FILING DATE: 13-AUG-1997
30     : CLASSIFICATION: 435
31     :
32     : PRIOR APPLICATION DATA:
33     : APPLICATION NUMBER: JP 214509/1996
34     : FILING DATE: 14-AUG-1996
35     :
36     : PRIOR APPLICATION DATA:
37     : APPLICATION NUMBER: JP 199979/1997
38     : FILING DATE: 25-JUL-1997
39     : ATTORNEY/AGENT INFORMATION:
40     : NAME: HALL, LISA A.
41     :
42     : REGISTRATION NUMBER: 38,347
43     : REFERENCE/DOCKET NUMBER: 07898/016001
44     :
45     : TELECOMMUNICATION INFORMATION:
46     : TELEPHONE: 619-678-5070
47     : TELEFAX: 619-678-5099
48     :
49     : INFORMATION FOR SEQ ID NO: 1:
50     : SEQUENCE CHARACTERISTICS:
51     : LENGTH: 1785 base pairs
52     : TYPE: nucleic acid
53     : STRANDEDNESS: double
54     : TOPOLOGY: linear
55     : MOLECULE TYPE: DNA (genomic)
56     :
57     : FEATURE:
58     : NAME/KEY: CDS
59     : LOCATION: 1..1782
60     :
61     :
62     :
63     :
64     :
65     :
66     :
67     :
68     :
69     :
70     :
71     :
72     :
73     :
74     :
75     :
76     :
77     :
78     :
79     :
80     :
81     :
82     :
83     :
84     :
85     :
86     :
87     :
88     :
89     :
90     :
91     :
92     :
93     :
94     :
95     :
96     :
97     :
98     :
99     :
100    :
101    :
102    :
103    :
104    :
105    :
106    :
107    :
108    :
109    :
110    :
111    :
112    :
113    :
114    :
115    :
116    :
117    :
118    :
119    :
120    :
121    :
122    :
123    :
124    :
125    :
126    :
127    :
128    :
129    :
130    :
131    :
132    :
133    :
134    :
135    :
136    :
137    :
138    :
139    :
140    :
141    :
142    :
143    :
144    :
145    :
146    :
147    :
148    :
149    :
150    :
151    :
152    :
153    :
154    :
155    :
156    :
157    :
158    :
159    :
160    :
161    :
162    :
163    :
164    :
165    :
166    :
167    :
168    :
169    :
170    :
171    :
172    :
173    :
174    :
175    :
176    :
177    :
178    :
179    :
180    :
181    :
182    :
183    :
184    :
185    :
186    :
187    :
188    :
189    :
190    :
191    :
192    :
193    :
194    :
195    :
196    :
197    :
198    :
199    :
200    :
201    :
202    :
203    :
204    :
205    :
206    :
207    :
208    :
209    :
210    :
211    :
212    :
213    :
214    :
215    :
216    :
217    :
218    :
219    :
220    :
221    :
222    :
223    :
224    :
225    :
226    :
227    :
228    :
229    :
230    :
231    :
232    :
233    :
234    :
235    :
236    :
237    :
238    :
239    :
240    :
241    :
242    :
243    :
244    :
245    :
246    :
247    :
248    :
249    :
250    :
251    :
252    :
253    :
254    :
255    :
256    :
257    :
258    :
259    :
260    :
261    :
262    :
263    :
264    :
265    :
266    :
267    :
268    :
269    :
270    :
271    :
272    :
273    :
274    :
275    :
276    :
277    :
278    :
279    :
280    :
281    :
282    :
283    :
284    :
285    :
286    :
287    :
288    :
289    :
290    :
291    :
292    :
293    :
294    :
295    :
296    :
297    :
298    :
299    :
300    :
301    :
302    :
303    :
304    :
305    :
306    :
307    :
308    :
309    :
310    :
311    :
312    :
313    :
314    :
315    :
316    :
317    :
318    :
319    :
320    :
321    :
322    :
323    :
324    :
325    :
326    :
327    :
328    :
329    :
330    :
331    :
332    :
333    :
334    :
335    :
336    :
337    :
338    :
339    :
340    :
341    :
342    :
343    :
344    :
345    :
346    :
347    :
348    :
349    :
350    :
351    :
352    :
353    :
354    :
355    :
356    :
357    :
358    :
359    :
360    :
361    :
362    :
363    :
364    :
365    :
366    :
367    :
368    :
369    :
370    :
371    :
372    :
373    :
374    :
375    :
376    :
377    :
378    :
379    :
380    :
381    :
382    :
383    :
384    :
385    :
386    :
387    :
388    :
389    :
390    :
391    :
392    :
393    :
394    :
395    :
396    :
397    :
398    :
399    :
400    :
401    :
402    :
403    :
404    :
405    :
406    :
407    :
408    :
409    :
410    :
411    :
412    :
413    :
414    :
415    :
416    :
417    :
418    :
419    :
420    :
421    :
422    :
423    :
424    :
425    :
426    :
427    :
428    :
429    :
430    :
431    :
432    :
433    :
434    :
435    :
436    :
437    :
438    :
439    :
440    :
441    :
442    :
443    :
444    :
445    :
446    :
447    :
448    :
449    :
450    :
451    :
452    :
453    :
454    :
455    :
456    :
457    :
458    :
459    :
460    :
461    :
462    :
463    :
464    :
465    :
466    :
467    :
468    :
469    :
470    :
471    :
472    :
473    :
474    :
475    :
476    :
477    :
478    :
479    :
480    :
481    :
482    :
483    :
484    :
485    :
486    :
487    :
488    :
489    :
490    :
491    :
492    :
493    :
494    :
495    :
496    :
497    :
498    :
499    :
500    :
501    :
502    :
503    :
504    :
505    :
506    :
507    :
508    :
509    :
510    :
511    :
512    :
513    :
514    :
515    :
516    :
517    :
518    :
519    :
520    :
521    :
522    :
523    :
524    :
525    :
526    :
527    :
528    :
529    :
530    :
531    :
532    :
533    :
534    :
535    :
536    :
537    :
538    :
539    :
540    :
541    :
542    :
543    :
544    :
545    :
546    :
547    :
548    :
549    :
550    :
551    :
552    :
553    :
554    :
555    :
556    :
557    :
558    :
559    :
560    :
561    :
562    :
563    :
564    :
565    :
566    :
567    :
568    :
569    :
570    :
571    :
572    :
573    :
574    :
575    :
576    :
577    :
578    :
579    :
580    :
581    :
582    :
583    :
584    :
585    :
586    :
587    :
588    :
589    :
590    :
591    :
592    :
593    :
594    :
595    :
596    :
597    :
598    :
599    :
600    :
601    :
602    :
603    :
604    :
605    :
606    :
607    :
608    :
609    :
610    :
611    :
612    :
613    :
614    :
615    :
616    :
617    :
618    :
619    :
620    :
```

	Query Match	8.0%	Score 34.4	DB 2	Length 1785
	Best Local Similarity	47.6%	Freq. No. 1.4		
	Matches 101	Conservative	0	Mismatches 111	Indels 0
					Gaps 0
Qy	139	gagggcgagctacgagtgatcgcgcgagagctccgcggggtcgcacccacaagaagctccac	198		
Db	388	gagggcgctccccaagaagacccggagcgcgttcgctttttaccgccacagactcaaac	447		
Qy	199	attatgctcgcgatgctgctcagctgaacaaagccgcagccgacccgacgaagaagactta	258		

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 12, 2001, 20:07:53 ; Search time 244.98 Seconds

(without alignments)
662.447 Million cell updates/sec

Title: US-09-461-774-7

Perfect score: 432
Sequence: 1 gccaccaccccttcacctca.....agatccggtccacacactga 432

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 480022 seqs, 187831343 residues

tal number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_36: *
1: /SID56/gcgdata/geneseq/geneseqn/NA1980.DAT: *
2: /SID56/gcgdata/geneseq/geneseqn/NA1981.DAT: *
3: /SID56/gcgdata/geneseq/geneseqn/NA1982.DAT: *
4: /SID56/gcgdata/geneseq/geneseqn/NA1983.DAT: *
5: /SID56/gcgdata/geneseq/geneseqn/NA1984.DAT: *
6: /SID56/gcgdata/geneseq/geneseqn/NA1985.DAT: *
7: /SID56/gcgdata/geneseq/geneseqn/NA1986.DAT: *
8: /SID56/gcgdata/geneseq/geneseqn/NA1987.DAT: *
9: /SID56/gcgdata/geneseq/geneseqn/NA1988.DAT: *
10: /SID56/gcgdata/geneseq/geneseqn/NA1989.DAT: *
11: /SID56/gcgdata/geneseq/geneseqn/NA1990.DAT: *
12: /SID56/gcgdata/geneseq/geneseqn/NA1991.DAT: *
13: /SID56/gcgdata/geneseq/geneseqn/NA1992.DAT: *
14: /SID56/gcgdata/geneseq/geneseqn/NA1993.DAT: *
15: /SID56/gcgdata/geneseq/geneseqn/NA1994.DAT: *
16: /SID56/gcgdata/geneseq/geneseqn/NA1995.DAT: *
17: /SID56/gcgdata/geneseq/geneseqn/NA1996.DAT: *
18: /SID56/gcgdata/geneseq/geneseqn/NA1997.DAT: *
19: /SID56/gcgdata/geneseq/geneseqn/NA1998.DAT: *
20: /SID56/gcgdata/geneseq/geneseqn/NA1999.DAT: *
21: /SID56/gcgdata/geneseq/geneseqn/NA2000.DAT: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	428.8	99.3	474	20	X17761
2	425.8	98.6	1081	20	Z20195
3	425.8	98.6	1104	20	Z20204
4	425.8	98.6	1299	20	Z20200
5	425.8	98.6	2577	20	Z20199
6	40.6	9.2	2742	19	V49291
7	39.8	9.2	1108	14	Q49396
8	39.4	9.1	972	16	T00591
9	39.4	9.1	2672	16	T08693
10	39	9.0	4897	11	O03259
11	37	8.6	707	20	Z16554
12	37	8.6	1894	18	T66542

13	36.4	8.4	2760	16	Q79921	P. aeruginosa accb
14	36.2	8.4	8625	18	T45143	Flavobacterium car
15	36.2	8.4	8625	19	V40146	Flavobacterium sp.
16	36.2	8.4	11233	19	V40151	DNA sequence of pl
17	36.2	8.4	17091	20	X60096	Acetobacter xylinu
18	36	8.3	351	21	A31539	Plant microsateili
19	36	8.3	1800	15	O70237	P. aeruginosa muca
20	36	8.3	1800	15	O70238	P. aeruginosa muca
21	35.2	8.1	1109	18	T91523	Mycobacterium tube
22	35.2	8.1	1109	18	T91457	Mycobacterium tube
23	35.2	8.1	1538	18	T63302	A. chrysogenum O-a
24	35.2	8.1	2634	14	Q39093	Streptomyces nodos
25	35.2	8.1	2860	18	T63303	A. chrysogenum O-a
26	35.2	8.1	3027	19	V64506	M. tuberculosis tm
27	35.2	8.1	3027	19	V44397	Mycobacterium tube
28	35.2	8.1	3027	20	Z19307	M. tuberculosis an
29	35.2	8.1	3027	20	Z19095	M. tuberculosis re
30	35	8.1	7926	21	Z89190	Plasmid TT01A 103L
31	34.8	8.1	3060	21	D00333	Rice raffinose syn
32	34.4	8.0	1785	19	V15493	Aeromonas caviae p
33	34.4	8.0	1785	20	Z31771	Polyester synthase
34	34.4	8.0	1785	20	Z25625	Aeromonas caviae p
35	34.4	8.0	3187	19	V15499	Aeromonas caviae p
36	34.4	8.0	3187	19	V15498	Aeromonas caviae p
37	34.4	8.0	3187	20	Z31773	Full length polyes
38	34.4	8.0	3187	20	Z25628	A. caviae polyeste
39	34.4	8.0	34094	20	Z20163	Complete nucleoid
40	34.2	7.9	4104	21	Z65053	Membrane-bound pro
41	34	7.9	2888	17	T59268	Streptomyces prist
42	33.8	7.8	352	21	A31634	Plant microsateili
43	33.8	7.8	371	21	A31988	Plant microsateili
44	33.8	7.8	620	17	T29126	TP08 gene fragment
45	33.8	7.8	620	17	T29124	TP08 gene fragment

ALIGNMENTS

RESULT 1	
ID X17761	standard; DNA; 474 BP.
XX	
AC X17761:	
DT 11-MAY-1999	(first entry)
XX	
DE	Acr gene encoding alpha-crystallin heat shock protein.
XX	
KW	Tuberculosis; Mycobacterium; alpha-crystallin heat shock protein;
KW	acr gene; attenuated; vaccine; antimycobacterial; pathogenic; ss.
XX	
OS	Mycobacterium sp.
XX	
FH	Key
FT	CDS
FT	Location/Qualifiers
XX	/*tag= a
XX	
PN	W09902670-A1.
PD	21-JAN-1999.
XX	
PF	09-JUL-1998; 98WO-US14227.
XX	
PR	10-JUL-1997; 97US-0052199.
XX	
PA	(USSH) US DEPT HEALTH & HUMAN SERVICES.
XX	
PI	Barry CE, Crane D, Yuan Y;
XX	
XX	WPI, 1999-120871/10.
DR	P-PSDB; W94900.
XX	
PT	Novel attenuated strains of Mycobacterium tuberculosis and M. bovis

PT - useful in the production of a vaccine against tuberculosis,
 PT achieved by reducing or eliminating expression of the
 PT alpha-crystallin heat shock protein gene

PS Examples: Page 25; 33pp; English.

XX The invention relates to a vaccine for protection against tuberculosis in
 CC which *Mycobacterium* sp. are attenuated by having the expression of the
 CC alpha-crystallin heat shock protein gene (*acr* gene) reduced by at least
 CC 75 percent. The attenuated strains can be used as models for screening
 CC for novel antimycobacterial agents. Also, the mixture of *Mycobacterium*
 CC sp. cellular debris can be used for detection of exposure to *M. bovis*
 CC and *M. tuberculosis*, used for commercial adjuvants, used as diagnostics
 CC including use as reagents for the development of monoclonal antibodies
 CC to recognize wild type *Mycobacterium* in patient samples. The vaccine is
 CC based on *M. tuberculosis* and is obtained without the multitude of
 CC passages so that the immune response generated is more relevant to the
 CC wild type, pathogenic bacteria. The present sequence represents the *acr*
 CC gene encoding the alpha-crystallin heat shock protein.

Sequence 474 BP; 94 A; 148 C; 142 G; 90 T; 0 other;

Query Match 99.3%; Score 428.8; DB 20; Length 474;
 Best Local Similarity 99.5%; Pred. No. 7.5e-105;
 Matches 430; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 gccacacacccctccggttcagcgccacccggtcccttcccccgggttttcgaactg 60
 DB 19 gccacacacccctccggttcagcgccacccggtcccttcccccgggttttcgaactg 78
 OY 61 ttcggagcctcccggttcagcgccacccggtcccttcccccgggttttcgaactg 120
 DB 79 ttcggagcctcccggttcagcgccacccggtcccttcccccgggttttcgaactg 138
 OY 121 ctggaaagacagatgaagaaggggcgctacgaaggtacgcgcggaagctcccggttcgac 180
 DB 139 ctggaaagacagatgaagaaggggcgctacgaaggtacgcgcggaagctcccggttcgac 198
 OY 181 cccgacaagagacgtccacatattgctcgacatgctacgtacacatcaaaagccgagcgc 240
 DB 199 cccgacaagagacgtccacatattgctcgacatgctacgtacacatcaaaagccgagcgc 258
 OY 241 accgagcagaagacatagaacggtcgctcggaattcggttcccttcgttcgacg 300
 DB 259 accgagcagaagacatagaacggtcgctcggaattcggttcccttcgttcgacg 318
 OY 301 gttctgctgcggtgaagtgctgaagagacatataagccacctcgaagaaggcatt 360
 DB 319 gttctgctgcggtgaagtgctgaagagacatataagccacctcgaagaaggcatt 378
 OY 361 cttaactgtcggtgaggttttcggaagggaagcacaacgaaagacattcaattccgg 420
 DB 379 cttaactgtcggtgaggttttcggaagggaagcacaacgaaagacattcaattccgg 438
 OY 421 tcacacacactga 432
 DB 439 tcacacacactga 450

RESULT 2
 ID 220195
 XX 220195 standard; DNA; 1081 BP.

AC 220195;

XX 17-JAN-2000 (first entry)

DE *Mycobacterium tuberculosis* antigen fusion protein Mtb39A DNA.

KM Tuberculosis; antigen; fusion protein; Mtb39A; EMD14; DPV; MTI;
 KW diagnosis; therapy; vaccine; immunogen; ss.

OS *Mycobacterium tuberculosis*.

XX Key Location/Qualifiers

FT CDS 10..1002

XX /*tag= a

PN W09951748-A2.

PD 14-OCT-1999.

XX 07-APR-1999; 99MO-US07717.

XX 07-APR-1998; 98US-0056556.

XX 30-DEC-1998; 98US-0223040.

PA (CORI-) CORIXA CORP.

PI Skeiky YAW, Alderson M, Campos-Neto A;

DR WPI; 1999-601610/51.

XX P-PSDB; Y32060.

PT New fusion proteins useful for diagnosis, prevention and treatment of

PT tuberculosis -

XX Example; Fig 2; 83pp; English.

CC This DNA sequence includes a coding region for a recombinant
 CC *Mycobacterium tuberculosis* tri-antigen fusion protein (see Y32060),
 CC termed Mtb39A, composed of the antigens EMD14, DPV and MTI. The
 CC DNA is useful for the recombinant production of the fusion protein.
 CC Coding sequences for the antigen were modified by PCR in order
 CC to facilitate their fusion and subsequent expression of the fusion
 CC protein. 3 Coding sequences for EMD14, DPV and MTI were ligated
 CC to encode Mtb39A. The invention provides fusion proteins (see
 CC Y32059-71) containing at least 2 *M. tuberculosis* antigens. The new
 CC fusion proteins and polynucleotides encoding them are useful as
 CC vaccines for preventing tuberculosis (claimed), for diagnosis (via
 CC in vitro assays or intradermal skin tests for detection of anti-M.
 CC tuberculosis antibodies), monitoring of disease progression, and
 CC treatment of tuberculosis. They are more effective immunogens than
 CC mixtures of the individual protein components.

SQ Sequence 1081 BP; 209 A; 341 C; 325 G; 206 T; 0 other;

Query Match 98.6%; Score 425.8; DB 20; Length 1081;
 Best Local Similarity 99.5%; Pred. No. 5.6e-104;
 Matches 427; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 gccacacacccctccggttcagcgccacccggtcccttcccccgggttttcgaactg 60
 DB 34 gccacacacccctccggttcagcgccacccggtcccttcccccgggttttcgaactg 93
 OY 61 ttcggagcctcccggttcagcgccacccggtcccttcccccgggttttcgaactg 120
 DB 94 ttcggagcctcccggttcagcgccacccggtcccttcccccgggttttcgaactg 153
 OY 121 ctggaaagacagatgaagaaggggcgctacgaaggtacgcgcggaagctcccggttcgac 180
 DB 154 ctggaaagacagatgaagaaggggcgctacgaaggtacgcgcggaagctcccggttcgac 213
 OY 181 cccgacaagagacgtccacatattgctcgacatgctacgtacacatcaaaagccgagcgc 240
 DB 214 cccgacaagagacgtccacatattgctcgacatgctacgtacacatcaaaagccgagcgc 273
 OY 241 accgagcagaagacatagaacggtcgctcggaattcggttcccttcgttcgacg 300
 DB 274 accgagcagaagacatagaacggtcgctcggaattcggttcccttcgttcgacg 333
 OY 301 gttctgctgcggtgaagtgctgaagagacatataagccacctcgaagaaggcatt 360
 DB 334 gttctgctgcggtgaagtgctgaagagacatataagccacctcgaagaaggcatt 393

Oy	61	ttcgcgagctctccggtcatcttcgcgggaactccgcccacattcgacacccggttgatgcyg	120
Oy	61	ttcgcgagctctccggtcatcttcgcgggaactccgcccacattcgacacccggttgatgcyg	120
Db	88	ttcgcgagctctccggtcatcttcgcgggaactccgcccacattcgacacccggttgatgcyg	147
Oy	121	ctggaagacgagatgaaagaggcgctacgaggttacgcgcggagctctccggggctgac	180
Db	148	ctggaagacgagatgaaagaggcgctacgaggttacgcgcggagctctccggggctgac	207
Oy	181	cccgacaagagacgtccacattatgtccgcgatagtctagcttgacacatcaagccgagcgc	240
Db	208	cccgacaagagacgctccacattatgtctagcttgacacatcaagccgagcgc	267
Oy	241	accgagcagaagagacttagacgctgcctcggaattcgctgacggttccctcttcgacg	300
Db	268	accgagcagaagagacttagacgctgcctcggaattcgctgacggttccctcttcgacg	327
Oy	301	gtgtcgctgcgggttagtggtcgtgaagagagacacttaagccacctacgacaaggagcat	360
Db	328	gtgtcgctgcgggttagtggtcgtgaagagagacacttaagccacctacgacaaggagcat	387
Oy	361	cttaactgtctgcgtgagcggtttcggaagggagaccacgcaaaagacattcagatccg	420
Db	388	cttaactgtctgcgtgagcggtttcggaagggagaccacgcaaaagacattcagatccg	447
Oy	421	tccacccaac 429	
Db	448	tccacccaac 456	
RESULT	4		
ID	220200		
XX	220200	standard; DNA; 1299 BP.	
AC	220200:		
XX	17-JAN-2000	(first entry)	
DT	17-JAN-2000	(first entry)	
XX			
DE		Mycobacterium tuberculosis antigen fusion protein Mtb46f DNA.	
XX			
KM		Tuberculosis: antigen; fusion protein; Mtb46f; EMD14; DPV; MTI; MSJ; diagnosis; therapy; vaccine; immunogen; ds.	
XX			
OS		Mycobacterium tuberculosis.	
XX			
PN		W09951748-A2.	
XX		14-OCT-1999.	
PD			
XX			
PF		07-APR-1999; 99WO-US07717.	
XX			
PR		07-APR-1998; 98US-0056556.	
PR		30-DEC-1998; 98US-0223040.	
XX			
PA		(CORI-) CORIXA CORP.	
XX			
PI		Skelky YAW, Alderson M, Campos-Neto A;	
XX		WPI: 1999-601610/51.	
DR		P-PSDB: Y32065.	
XX			
PT		New fusion proteins useful for diagnosis, prevention and treatment of tuberculosis -	
XX			
PS		Example; Fig 7A-B; 83pp; English.	
XX			
CC	This DNA sequence represents a coding region for a recombinant		
CC	Mycobacterium tuberculosis tetra-antigen fusion protein (see Y32065),		
CC	termed Mtb46f, composed of the antigens ERD14, DPV, MTI and MSJ.		
CC	The DNA is useful for the recombinant production of Mtb46f. The		
CC	coding sequences for the antigens were modified by PCR in order to		
CC	facilitate their fusion and subsequent expression of the fusion		
CC	protein. The coding sequences for the 4 antigens were ligated		
CC	to encode Mtb46f. The invention provides fusion proteins (see		

CC Y32059-71) containing at least 2 M. tuberculosis antigens. The new
CC fusion proteins and polynucleotides encoding them are useful as
CC vaccines for preventing tuberculosis (claimed), for diagnosis (via
CC in vitro assays or intradermal skin tests for detection of anti-M.
CC tuberculosis antibodies), monitoring of disease progression, and
CC treatment of tuberculosis. They are more effective immunogens than
CC mixtures of the individual protein components.
SQ Sequence 1299 BP; 234 A; 407 C; 408 G; 250 T; 0 other;

Query Match 98.6%; Score 425.8; DB 20; Length 1299;
Best Local Similarity 99.5%; Pred. No. 5.8e-104;
Matches 427; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gccacacacccctccggttaagcgccaccccggttcctctcccgagtttctgagctg 60
Db 28 gccacacacccctccggttaagcgccaccccggttcctctcccgagtttctgagctg 87

61 ttccgagcctcccggttaagcgccaccccggttcctctcccgagtttctgagctg 120
88 ttccgagcctcccggttaagcgccaccccggttcctctcccgagtttctgagctg 147

QY 121 ctggaagacgagatgaagaaggcgctacagagttacgcgagagcttcccggttcgac 180
Db 148 ctggaagacgagatgaagaaggcgctacagagttacgcgagagcttcccggttcgac 207

QY 181 ccggaagaagagctccacattatggtccgagatgcagttgacattcaaggcgagcgc 240
Db 208 ccggaagaagagctccacattatggtccgagatgcagttgacattcaaggcgagcgc 267

QY 241 accgagacaagagacttaagcgttcgctcggaattcgctacggttcccttcgttcgacg 300
Db 268 accgagacaagagacttaagcgttcgctcggaattcgctacggttcccttcgttcgacg 327

QY 301 ggttcgctccggttagtgcgtgaagagagacatataaggccaactacgaagaagcatt 360
Db 328 ggttcgctccggttagtgcgtgaagagagacatataaggccaactacgaagaagcatt 387

QY 361 cttaactgttcggttcggttttcggaagggaagcaacccaagaacatttcagtcgag 420
Db 388 cttaactgttcggttcggttttcggaagggaagcaacccaagaacatttcagtcgag 447

QY 421 tccaccaac 429
Db 448 tccaccaac 456

●
ID 220199 standard; DNA; 2577 BP.
AC 220199;

DT 17-JAN-2000 (first entry)
XX
DE Mycobacterium tuberculosis antigen fusion protein Mtb88f DNA.

XX
KM Tuberculosis; antigen: fusion protein; Mtb88f; ERD14; DPV; M71;
KW MSL; M7CC2; diagnosis; therapy; vaccine; immunogen; ds.

XX
OS Mycobacterium tuberculosis.
XX
FH Key Location/Qualifiers
FT CDS 1..2571
ET CDS /tag= a

PN WO951748-A2.
XX
PD 14-OCT-1999.
XX
PF 07-APR-1999; 99WO-US07717.
XX

PR 07-APR-1998; 9805-0056556.
PR 30-DEC-1998; 9805-0223040.
XX
PA (COR-) CORIXA CORP.
XX
PI Skeiky YAW, Alderson M, Campos-Neto A;
XX
DR WPI; 1999-601610/51.
DR P-PSDB; Y32064.
XX
PT New fusion proteins useful for diagnosis, prevention and treatment of
PT tuberculosis -
XX
XX Example; Fig 6A-B; 83pp; English.

CC This DNA sequence represents a coding region for a recombinant
CC Mycobacterium tuberculosis penta-antigen fusion protein (see Y32064),
CC termed Mtb88f, composed of the antigens ERD14, DPV, M71, MSL and
CC M7CC2. The DNA is useful for the recombinant production of Mtb88f.
CC Coding sequences for the antigens were modified by PCR in order
CC to facilitate their fusion and subsequent expression of the fusion
CC protein. The coding sequences for the 5 antigens were ligated
CC to encode Mtb88f. The invention provides fusion proteins (see
CC Y32059-71) containing at least 2 M. tuberculosis antigens. The new
CC fusion proteins and polynucleotides encoding them are useful as
CC vaccines for preventing tuberculosis (claimed), for diagnosis (via
CC in vitro assays or intradermal skin tests for detection of anti-M.
CC tuberculosis antibodies), monitoring of disease progression, and
CC treatment of tuberculosis. They are more effective immunogens than
CC mixtures of the individual protein components.
SQ Sequence 2577 BP; 441 A; 840 C; 861 G; 435 T; 0 other;

Query Match 98.6%; Score 425.8; DB 20; Length 2577;
Best Local Similarity 99.5%; Pred. No. 6.7e-104;
Matches 427; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gccacacacccctccggttaagcgccaccccggttcctctcccgagtttctgagctg 60
Db 28 gccacacacccctccggttaagcgccaccccggttcctctcccgagtttctgagctg 87

QY 61 ttccgagcctcccggttaagcgccaccccggttcctctcccgagtttctgagctg 120
Db 88 ttccgagcctcccggttaagcgccaccccggttcctctcccgagtttctgagctg 147

QY 121 ctggaagacgagatgaagaaggcgctacagagttacgcgagagcttcccggttcgac 180
Db 148 ctggaagacgagatgaagaaggcgctacagagttacgcgagagcttcccggttcgac 207

QY 181 ccggaagaagagctccacattatggtccgagatgcagttgacattcaaggccaaggcgagc 240
Db 208 ccggaagaagagctccacattatggtccgagatgcagttgacattcaaggccaaggcgagc 267

QY 241 accgagacaagagacttaagcgttcgctcggaattcgctacggttcccttcgttcgacg 300
Db 268 accgagacaagagacttaagcgttcgctcggaattcgctacggttcccttcgttcgacg 327

QY 301 ggttcgctccggttagtgcgtgaagagagacatataaggccaactacgaagaagcatt 360
Db 328 ggttcgctccggttagtgcgtgaagagagacatataaggccaactacgaagaagcatt 387

QY 361 cttaactgttcggttcggttttcggaagggaagcaacccaagaacatttcagtcgag 420
Db 388 cttaactgttcggttcggttttcggaagggaagcaacccaagaacatttcagtcgag 447

QY 421 tccaccaac 429
Db 448 tccaccaac 456

RESULT 6
VA9291

ID	V49291 standard; DNA; 2742 BP.
XX AC	V49291;
XX DT	28-OCT-1998 (first entry)
DE XX	Aujeszky's disease virus glycoprotein gB gene.
KW KW	Multivalent vaccine; p1g: pathogen; respiratory disease; STV; PRRSV; HCV; digestive disease; Aujeszky's disease virus; pseudorabies virus; vaccine; swid herpesvirus 1; swine influenza virus; hog cholera virus; vector;
KW KW	porcine respiratory and reproductive syndrome virus; glycoprotein; SIRSV; swine infertility and respiratory syndrome virus; <i>Actinobacillus pleuropneumoniae</i> ; ss.
XX OS	Pseudorabies virus.
XX FH	Key Location/Qualifiers
FT CDS	1..2742 /*tag= a /product= "glycoprotein gB"
XX PN	FR2751224-AI.
PD PD	23-JAN-1998.
XX PF	19-JUL-1996; 96FR-0009338.
XX PE	19-JUL-1996; 96FR-0009338.
PR PA	(INMR) RHONE MERIEUX SA. MPI; 1998-112824/I1. P-PDB; W68404.
DR DR	Multi-valent polynucleotide vaccines against porcine pathogens - consist of at least 3 plasmids able to express protective antigens from specified viruses
PT PS	Example 8; Fig 2; 63pp; French.
XX CC	The invention relates to a multivalent vaccine for protecting pigs against several pathogens, especially pathogens associated with respiratory and digestive diseases. The pathogens are especially selected from Aujeszky's disease virus, swine influenza virus (SIV), porcine respiratory and reproductive syndrome virus (RRSV), hog cholera virus (HCV) and <i>Actinobacillus pleuropneumoniae</i> . The vaccines are preferably composed of polynucleotide sequences encoding 3 antigens, all as part of vectors. This sequence represents the coding region for the Aujeszky's disease virus strain NIA3 glycoprotein gB. The sequence was subcloned into the plasmid pVR1012 to generate plasmid pAB090 for use in the vaccine.
CC CC	
XX CC	
SO	Sequence 2742 BP; 416 A; 1024 C; 928 G; 374 T; 0 other;
QY Query Match	9.4%; Score 40.6; DB 19; Length 2742; Best Local Similarity 49.3%; Pred. No. 0.063; Mismatches 109; Indels 0; Gaps 0 Matches 106; Conservative 0;
Db 41	tcccccagattttcttgaagtgttcgcggactcccgatcatgccgcggactccgccccacct 100
Db 2414	tgttgttgctggtcgccgctgcgtgcggcttctctgaccataccgatcatcgtgcgttcg 2473
QY 101	tcgacaaccgcgttgcgtgcgtgcgaagcacgatatgaaaggaggcgctacgagltacgcg 160
Db 2474	gocgcaaaccccatgaaggccctgtaccctccaagaagcagcgtcaaaggagacgcg 2533
QY 161	cggaaacttcccgaggttcgaaccccgcaagaagacgttcaccaattatgttcgcgagtgcacg 220
Db 2534	tcgacaaggcgcaagctvgagagagccaaaagctvgaaacgaagcccgagacatgatlccggtta 2593
QY 221	tgaccataaaggcgcgagcgaccacgcgagcaaaagac 255

[illegible]

Query Match	Best Local Similarity	Score	DB	Length	972:
Matches	121;	Conservative	0;	Mismatches	136;
				Indels	0;
				Gaps	0;
QY	120	gctgaagacgaagatgaagagggcgctacgaaggtacgcgcggagctccgggggtcga	179		
DB	45	ggccgacgaagagaggtccgcgaggtggccgcgcggcggtcttctaccctcaccgggtacgg	104		
QY	180	cccgcgaagaagagcgtccacattatggtccgcgatatggtcgcgtgacacataagcgccagcg	239		

Db	105	cgccaccgaagaagagaccacccgggtggccaccgacacccgcga tggactcttcgcccaagg	164
Qy	240	caccgacgaagaaggactaagacggtcgcgtcggaattcgctagcgttccttcgtcgac	298
Db	165	cacgagccgagagaagaagcagcgcttgaccacgaaggtcccgaccatgcgcgcggtactc	224
Qy	300	ggctgcgtgcgcggtaggctgctgacgaggaacatlaagccacactagacaaggcat	355
Db	225	ggcgtctgagagcggaagacacccgcccaggctacacaacccgcacactacacgactatc	284
Qy	360	tcttactgtgcggtg 376	
Db	285	catgtcgtacctcgatg 301	
RESULT	9		
ID	T08693		
AC	T08693	standard; cDNA; 2672 BP.	
XX			
DT	25-AUG-1996	(first entry)	
XX			
DE		Cephamecin biosynthetic enzyme gene cluster.	
XX			
KW		Cephamecin: cephalosporin; antibiotic;	
KW		C-7 hydroxycephem methyltransferase; 3'-methylcephem hydroxylase;	
KW		3'-hydroxymethylcephem O-carbamoyltransferase; ss.	
XX			
OS		Nocardia lactamdurans strain LC411.	
XX			
FH	Key	Location/Qualifiers	
FT	CDS	1..70	
FT		/*tag= a	
FT		/label= pcbC	
FT		/note= "3' region of pcbC gene"	
FT	RBS	73..80	
FT		/*tag= b	
FT	CDS	84..794	
FT		/*tag= c	
FT		/label= ORF7	
FT		/product= C-7 hydroxycephem methyltransferase	
FT		/note= "(claim 8, page 40)"	
FT	misc_difference	504..602	
FT		/*tag= d	
FT		/note= "translated sequence of bases 503-602 given in the specification do not correspond to the Genetic Code"	
FT	RBS	785..789	
FT		/*tag= e	
FT		/note= "putative RBS for ORF7/ORF8 cotranslation"	
FT	CDS	802..1668	
FT		/*tag= f	
FT		/label= ORF8	
FT		/product= coupling protein	
FT		/note= "(Claim 7, page 39-40)"	
FT	RBS	1678..1685	
FT		/*tag= g	
FT	CDS	1692..2624	
FT		/*tag= h	
FT		/label= ORF9	
FT		/product= 3'-methylcephem-hydroxylase	
FT		/note= "ceif gene (Claim 5, page 39)"	
FT	stem_loop	2634..2667	
FT		/*tag= i	
FT		/note= "stem-loop may correspond to a transcription terminator"	
XX			
PN	W09529253-A1		
XX			
PD	02-NOV-1995.		
XX			
PF	17-APR-1995;	95WO-US04801.	

```
XX 22-APR-1994; 94US-0233605.
PR
XX
XX (MERI ) MERCK & CO INC.
PA
XX Coque JFR, Enguila EJ, Fuente JL, Liras P, Llaraena EJ;
PI Martin JF;
XX
DR WPI: 1995-382998/49.
DR P-PSDB; R92151, R92152, R92153, R92154.
XX
PT DNA encoding late enzymes involved in cephamycin biosynthesis - and
PT related proteins, used to transform cells for cephamycin prodn.
XX
PS Disclosure: Fig 3A-C: 61pp: English.
XX
CC An internal sequence (T08693) of a 5.4 kb BamHI cDNA fragment of the
CC Nocardia lactamdurans cephamycin C gene cluster was isolated from
CC a phage clone using a cefc gene probe. It contains the 3'
CC region of the pcbC gene, open reading frames (ORF7 and ORF8) that
CC code for a 2-component system comprising C-7 hydroxylase
CC methyltransferase (R92152) and a coupling protein (R92153), and the
CC cefr gene (ORF9) encoding 3'-methylcephem hydroxylase (R92154).
CC Another isolated cDNA (T08694) included the downstream cmcH gene
CC (ORF10) coding for 3'-hydroxymethylcephem O-carbamoyltransferase
CC (R92156). These sequences can be used for prodn. of recombinant
CC enzymes useful for industrial-scale cephamycin synthesis.
XX
SQ Sequence 2672 BP; 441 A; 952 C; 901 G; 378 T; 0 other;

Query Match          9.1%; Score 39.4; DB 16; Length 2672;
Best Local Similarity 47.1%; Pred. No. 0.13;
Matches 121; Conservative 0; Mismatches 136; Indels 0; Gaps 0;

QY 120 gctggaacagcagatgaagaaggcgctacgaagtaacgcggagcctcccggggtcga 179
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1745 ggcgcagcagcaggttcgcgagtggtgcccgcggggtctctctccacacgggtacgg 1804
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 180 cccgcacagaagcgtccacattatgttcgcgagatgtcaagctgaccatcaagcgcgagcg 239
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1805 cgcaccacgaagcagccacccgggtgtgccaacgcacgcgagtgactcttcgcgccaag 1864
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 240 caccgacgaagaaggacttagcgtgcgtcggaattcgatcgtagcgttcttcgttcgcac 299
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1865 cagcgcgaagagaagcagcgccgtgacacgaagaagttcccgaccatcgcgcggttactc 1924
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    300 ggtgcgtctccgtaggtgtgtacgaagacacattaaagccaccctacgaacagggcat 359
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    1925 ggcgtctgagggaggaagacacgcgcccaggttcaaccacacgcgacactacacgcgactactc 1984
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 360 tcttactgtgtcgtgtgg 376
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1985 catgtcgtactcgtatcg 2001
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 10
Q03259
ID Q03259 standard; DNA; 4897 BP.
XX
XX Q03259;
AC
XX 22-JUL-1990 (first entry)
DE
XX Pseudorabies virus glycoprotein gII polypeptide.
XX
XX Pseudorabies virus; vaccine; herpes virus; pig; ss.
XX
XX Pseudorabies virus.
OS
XX
XX Key Location/Qualifiers
FH CDS 2639..3001
FT /*tag= a
```

```
FT /note="gII glycoprotein"
XX
XX EP353809-A.
XX
XX 07-FEB-1990.
PD
XX
XX 18-JUL-1989; 89EP-0201887.
PF
XX
XX 01-AUG-1988; 88EP-0112479.
PR
XX
XX (ALKU ) AKZO NV.
PA
XX
XX Schreurs CS, Mettenleiter TC, Simon AJ, Lukacs N, Rhiza HJ;
PI
XX WPI: 1990-038534/06.
DR
XX P-PSDB; R03732.
XX
PT Immunogenic polypeptide of pseudo-rabies virus -
PT for use in mammalian vaccine against Aujeszky's disease.
XX
PS Disclosure: Fig 1; 21pp: English.
XX
CC The polypeptide may be used as a vaccine to provide
CC mammals, e.g. pigs, with resistance against Aujeszky's
CC disease.
XX
SQ Sequence 4897 BP; 594 A; 2052 C; 1495 G; 756 T; 0 other;

Query Match          9.0%; Score 39; DB 11; Length 4897;
Best Local Similarity 48.8%; Pred. No. 0.19;
Matches 105; Conservative 0; Mismatches 110; Indels 0; Gaps 0;

QY 41 tcccagatttctgagctgttcgcgacctccgcatctcgcgagactccgcccacct 100
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3285 tgtgtgtgtgcgcgcgctgttcgcgacctcttcctaccgacatctcgcgacctcgc 3344
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 101 tcgaaccccggttgatgcgctggaagacagatgaagaagggtcgtacaggtacg 160
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3345 gtccgaaccccatgaagccctgtacctaccgacgaagacagtgaggaagcg 3404
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 161 cggagcttcccggggtcgaccccgacgaagcagctccacattatgttccgcatgtcagc 220
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3405 tcgaagagggtcagctgtgacgaagcgaagctggaccaggcccgagacatgatccgltaca 3464
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 221 tgaccatcaaggcgcagcgacacgagcagaagagac 255
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3465 tgtcatcgtgttcgacctctgacgacagagagac 3499
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 11
Z16654
ID Z16654 standard; cDNA; 707 BP.
XX
XX Z16654;
AC
XX 12-OCT-1999 (first entry)
DE
XX Human gene expression product cDNA sequence SEQ ID NO:4124.
XX
XX Human: gene; gene expression product; diagnosis; therapy; probc;
XX detection; mapping; tissue typing; profiling; forensic; cancer;
XX genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.
XX
XX Homo sapiens.
OS
XX
XX W09938972-A2.
XX
XX 05-AUG-1999.
PD
XX
XX 28-JAN-1999; 99WO-US01619.
PF
XX
XX 03-APR-1998; 98US-0080666.
```



```

Db 1223 tgcagagcagcaagtgatcagctgctacgctgcgcagcgccggcctcccccagaccg 1282
QY 179 accccgacaaagacgtccacattatggtccgcgagtgatgacctcaagccgagc 238
Db 1283 actcgcctcaacgaggtctcgcgcgcacgagcgcgccgctacacctcgcagctcgcg 1342
QY 239 gcaccgacgagagac 255
Db 1343 gcaccgagcggtcgtggc 1359

RESULT 13
Q79921
ID Q79921 standard; DNA: 2760 BP.
AC Q79921;
XX 16-JUL-1995 (first entry)
DT P. aeruginosa accb and accc genes.
XX
XX Acetyl-CoA-carboxylase; accb gene; accc gene; malonyl-CoA; lipid;
KM oilseed; transgenic plant; biotin-carboxylase; BC;
KM biotin carboxyl carrier protein; BCCP; ss.
XX
OS Pseudomonas aeruginosa PAO.
XX
XX Key Location/Qualifiers
FH CDS 861..1334
FT /*tag= a
FT /label= accb
FT /note= "encodes the biotin carboxyl carrier
FT preprotein"
FT CDS 1349..2701
FT /*tag= b
FT /label= accc
FT /note= "encodes biotin-carboxylase"
XX
XX W09429467-A.
XX
XX PD 22-DEC-1994.
XX
XX PF 06-JUN-1994; 94WO-US06447.
XX
XX PR 08-JUN-1993; 93US-0074121.
XX
XX PA (CALJ ) CALGENE INC.
XX
XX Best E, Knauf V;
XX
XX WPI; 1995-036492/05.
XX DR P-PSDB; R66743; R66744.
XX
XX Increasing lipid levels in plants by modulating malonyl-CoA
XX levels - partic. for oilseed crops, also nucleic acid
XX constructs for expressing acetyl-CoA carboxylase, transgenic
XX plants and peptide prods.
XX
XX PS Disclosure; Fig. 4A-E; 48pp; English.
XX
XX CC The accb and accc genes of Pseudomonas aeruginosa PAO encode the
XX biotin carboxyl carrier protein (BCCP, given in R66743) and
XX biotin-carboxylase (BC, given in R66744) components, respectively,
XX of acetyl-CoA-carboxylase. The genes may be used to modulate lipid
XX accumulation in transgenic plants.
XX
XX SQ Sequence 2760 BP; 530 A; 962 C; 835 G; 433 T; 0 other:

Query Match 8.4%; Score 36.4; DB 16; Length 2760;
Best Local Similarity 50.9%; Pred. No. 0.82;
Matches 113; Conservative 0; Mismatches 106; Indels 3; Gaps 1;

```

```

QY 1 gccacacccctccgcttaacgcacccgcggtccctctcccccaggtttctgna-ctg 60
Db 1826 gccgcgcggcgcgcgcggtggcgcgcatgctgcgtggtctaacagagtcagactgac 1885
QY 61 ttgcgggcctccctcaltcgcgcgaactcgcgcacaccttcgaacaccggtlga-ctcg 120
Db 1886 aagtcgcccgaactgaccgcacgcgagcgcgcggtctcgcaaccg-lygtctac 1945
QY 121 ctggaagacgagatgaagaaggcgctcagagtgatcgcgcgagacttccgggtcgac 180
Db 1946 ctggaagagttcctcgtacacaccgcgcacgtggaagtccaggtgcttccga---cgcg 2002
QY 181 cccgacaagacgtccacattatggtccgcggtgctcagctg 222
Db 2003 cagggcaacgcacatccaccctcgcgcagcgcgactgctccctg 2044

RESULT 14
T45143/c
ID T45143 standard; DNA: 8625 BP.
XX
XX T45143;
XX
XX 08-MAR-1997 (first entry)
DT
XX
XX Flavobacterium carotenoid biosynthesis cluster DNA.
XX
XX Carotenoid; lycopene; beta-carotene; echinenone; canthaxanthin;
KM zeaxanthin; adonixanthin; astaxanthin; crtE;
KM geranylgeranyl pyrophosphate synthase; GGPP synthase; crtB;
KM prephytoene synthase; phytoene synthase; crtI; phytoene desaturase;
KM crtI; lycopene cyclase; crtZ; beta-carotene hydroxylase; ds.
XX
XX Flavobacterium sp. R1534 WT (ATCC 21588) .
XX
XX OS
XX FH Key Location/Qualifiers
XX FT CDS 2..1168
XX FT /*tag= a
XX FT /label= ORF-5
XX FT 1180..2355
XX FT /*tag= b
XX FT /label= ORF-1
XX FT 2521..3408
XX FT /*tag= c
XX FT /label= crtB
XX FT /product= (pre)phytoene synthase
XX FT complement (4313..5797)
XX FT /*tag= e
XX FT /label= crtI
XX FT /product= phytoene desaturase
XX FT complement (5794..6942)
XX FT /*tag= f
XX FT /label= crtY
XX FT /product= lycopene cyclase
XX FT complement (6939..7448)
XX FT /*tag= g
XX FT /label= crtZ
XX FT /product= beta-carotene hydroxylase
XX FT complement (7767..8315)
XX FT /*tag= h
XX FT /label= ORF-16
XX FT 8348..8349
XX FT misc_difference 8348-8349
XX FT /*tag= i
XX FT /note= "bases 8348-8349 are given as nn in the
XX specification"
XX
XX misc_difference 8539-8540
XX FT /*tag= j
XX FT /note= "bases 8539-8540 are given as nn in the
XX specification"
XX

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 12, 2001, 20:06:54 : Search time 6876.06 seconds
(without alignments)
321.331 Million cell updates/sec

Title: US-09-461-774-7
Perfect score: 432
Sequence: 1 gccaccacccttcctcgta.....agatccgcgtccaccactga 432

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 118133 seqs, 2558875100 residues
tal number of hits satisfying chosen parameters: 2236266

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_om:*
4: gb_ov:*
5: gb_ph:*
6: gb_pl1:*
7: gb_pl2:*
8: gb_pr1:*
9: gb_pr2:*
10: gb_pr3:*
11: gb_ro:*
12: gb_sy:*
13: gb_un:*
14: em_fun:*
15: em_hum1:*
16: em_hum2:*
17: em_in:*
18: em_om:*
19: em_or:*
20: em_ov:*
21: em_pat:*
22: em_ph:*
23: em_pl:*
24: em_ro:*
25: em_sts:*
26: em_sy:*
27: em_un:*
28: em_vi:*
29: gb_htg1:*
30: gb_htg2:*
31: gb_in1:*
32: gb_in2:*
33: em_ba1:*
34: em_ba2:*
35: em_hum3:*
36: em_hum4:*
37: gb_pr4:*
38: gb_htg3:*
39: gb_htg4:*
40: gb_htg5:*
41: gb_htg6:*
42: gb_htg7:*
43: em_htg1:*

44: em_htg2:*
45: em_htg3:*
46: em_hum5:*
47: gb_pl3:*
48: gb_pr5:*
49: gb_htg8:*
50: gb_htg9:*
51: gb_htg10:*
52: gb_htg11:*
53: gb_htg12:*
54: gb_htg13:*
55: gb_htg14:*
56: gb_in3:*
57: gb_htg15:*
58: gb_htg16:*
59: gb_htg17:*
60: em_htg4:*
61: em_htg5:*
62: em_htg6:*
63: em_htg7:*
64: em_hum6:*
65: gb_htg18:*
66: gb_htg19:*
67: gb_htg20:*
68: gb_htg21:*
69: gb_htg22:*
70: gb_htg23:*
71: gb_v11:*
72: gb_v12:*
73: gb_ba3:*
74: em_htg8:*
75: em_htg9:*
76: em_htg10:*
77: gb_pr6:*
78: gb_pr7:*
79: gb_sts1:*
80: gb_sts2:*
81: gb_dal1:*
82: gb_pat2:*
83: em_htg0:*
84: gb_htg24:*
85: gb_pr8:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	428.8	99.3	474	2 MSG1AKA	M76712 Mycobacteri
2	428.8	99.3	474	2 S79751	S79751.14 kda anti
3	428.8	99.3	53450	2 MTW018	AL021899 Mycobacte
4	48.2	11.2	528	1 AF072875	AF072875 Mycobacte
5	47.8	11.1	2982	1 AF134837	AF134837 Amycolato
6	47	10.9	22800	2 MTW034	AL021929 Mycobacte
7	44.6	10.3	43349	73 SCF41	AL117387 Streptomy
8	44.2	10.2	37521	73 SC1A8A	AL353861 Streptomy
9	44	10.2	31636	73 SC7C7	AL031031 Streptomy
10	43.8	10.1	110000	84 IMF1CHR12_14	Continuation (15 o
11	43.2	10.0	2748	73 SAUS1GA	M94370 Stigmatella
12	42.8	9.9	1750	1 AF257177	AF257177 Streptomy
13	42.2	9.8	399	72 SH027483	U27483 Suid herpes
14	42.2	9.8	2724	71 AF257079	AF257079 Pseudorab
15	42	9.7	3354	73 SRT15759	Y15759 Streptomyce
16	41.8	9.7	2747	47 NE0HSP30	M55672 Neurospora
17	41.8	9.7	3258	47 HAKOHE1B	D26472 Neurospora
18	41.6	9.6	1948	73 SPFAMAB	M61931 S.purpurasc
19	41.6	9.6	13086	73 SC1G2	AL136500 Streptomy
20	41.4	9.6	38366	73 SC23B6	AL390968 Streptomy
21	40.8	9.4	12021	1 AE004848	AE004848 Pseudomon

```

22 40.8 9.4 184457 1 AF079317 Sphingomo
23 40.6 9.4 2742 A68929 Sequence 1
24 40.6 9.4 3049 81 SHGP2SHV MI3321 Pseudorhabe
25 40.6 9.4 11767 1 AE004536 AE004536 Pseudomon
26 40.6 9.4 25797 73 SC2G58 AL391017 Streptomy
27 40.4 9.4 42210 73 SC1C2 AL031124 Streptomy
28 40.2 9.3 2760 34 SCASF8 D15062 Streptomyce
29 40.2 9.3 5955 73 STMAFSK1 D4382 Streptomyce
30 40.2 9.3 40969 73 SCDB6 AL353815 Streptomy
31 40 9.3 26500 73 SC9H11 AL356592 Streptomy
32 39.8 9.2 720 2 MKU20446 U20446 Mycobacteri
33 39.8 9.2 1546 73 SSPY15518 Y15518 Streptomyce
34 39.8 9.2 2261 47 ZMCABM7 X53398 Z.mays cab-
35 39.8 9.2 6162 1 AB016803 AB016803 Deinococc
36 39.8 9.2 10952 1 AE002014 AE002014 Deinococc
37 39.8 9.2 26440 73 SCE15 AL049707 Streptomy
38 39.8 9.2 184804 54 AC036174 AC036174 Homo sapi
39 39.6 9.2 41800 73 SCDBA AL160331 Streptomy
40 39.4 9.1 933 2 NLCEFA Z21687 N.lactamidur
41 39.4 9.1 41782 73 SCG11A AL133210 Streptomy
42 39.4 9.1 45335 73 SCC57A AL136519 Streptomy
43 39.2 9.1 918 6 AF089845 AF089845 Funaria h
44 39.2 9.1 42816 73 SCC88 AL139298 Streptomy
45 39.2 9.1 116868 30 AC009910 AC009910 Drosophila

```

ALIGNMENTS

RESULT 1

```

MSG14KA 474 bp DNA BCT 26-APR-1993
DEFINITION Mycobacterium tuberculosis 14 kD antigen gene, complete cds.
ACCESSION M76712.1 GI:149891
VERSION 14 kD antigen.
KEYWORDS 14 kD antigen.
SOURCE Mycobacterium tuberculosis (library: lambda-gl11; Lorisig:MTB) DNA.
ORGANISM Mycobacterium tuberculosis
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacteriinae; Mycobacteriaceae;
Mycobacterium; Mycobacterium tuberculosis complex.
1 (bases 1 to 474)
Verdon,A., Hartskeerl,R.A., Schultema,A., Kolk,A.H.J., Young,D.B.
and Lathigra,R.B.
The 14k antigen of Mycobacterium tuberculosis is related to. . .
Unpublished (1991)
FEATURES
    source          1..474
                     Location/Qualifiers
                     /organism="Mycobacterium tuberculosis"
                     /db_xref="taxon:1773"
                     /tissue_1lb="lambda-gl11; Lorisig:MTB"
                     /gene="14 kD antigen"
                     /note="14 kD antigen"
                     /codon_start=1
                     /transl_table=1
                     /product="14 kD antigen"
                     /protein_id="AAA25342.1"
                     /db_xref="GI:149892"
                     /translation="MATTLPVQRHRSLEPPESEFAAPPSFAGLRPTEDRLRMLED
EMKEGRYERVALEPVDKDVDMVRGQLTIKARETQKDFDGRSEFAVGSFVRTV
SLPVGADDDDKATYDKGILTVSVAVSGKPTKEKHQIRSTN"
BASE COUNT      94 a 148 c 142 g 90 t
ORIGIN

```

```

Query Match      99.3% Score 428.8; DB 2; Length 474;
Best Local Similarity 99.5%; Pred. No. 5,4e-68;
Matches 430; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
1 gccacacccttcgcgttaagcgccacccgggtcccttcccgagtttcagcg 60
|||||

```

```

Db 19 gccacacccttccggttacgcccacccgggtcccttccccgacgtttctgacgtc 78
Qy 61 ttcgcgacctcccgatcatccgcggaactccgcacacttcgacacccggttgcg 120
Db 79 ttgcggcgcttccgcatctccggagctccggacccacacttcgacacccgggttgcg 138
Qy 121 ctggaagacagatgaagaaggcgctacgaagtacgcgcggaacttccgggtgcac 180
Db 139 ctggaagacagatgaagaaggcgctacgaagtacgcgcggaacttccgggtgcac 198
Qy 181 cccgaagaagacgtccacattatgctccgcatgctcgaactgaacaaaggccggtcgc 240
Db 199 cccgaagaagacgtccacattatgctccgcatgctcgaactgaacaaaggccggtcgc 258
Qy 241 accgaagaagacgtccacattatgctccgcatgctcgaactgaacaaaggccggtcgc 300
Db 259 accgaagaagacgtccacattatgctccgcatgctcgaactgaacaaaggccggtcgc 318
Qy 301 gttgcgtccggttagtgcgtcgaagaacacattaaaggccacactgaagaaggcalt 360
Db 319 gttgcgtccggttagtgcgtcgaagaacacattaaaggccacactgaagaaggcalt 378
Qy 361 cttaactgttcggttcggttcggaagggaagcgaacgaagaacattgaatccg 420
Db 379 cttaactgttcggttcggttcggaagggaagcgaacgaagaacattgaatccg 438
Qy 421 tccacacactga 432
Db 439 tccacacactga 450

```

RESULT 2

```

LOCUS      579751 474 bp DNA BCT 07-MAY-1993
DEFINITION 14 kda antigen [Mycobacterium tuberculosis, Genomic, 474 nt].
ACCESSION 579751
VERSION 579751.1 GI:244561
KEYWORDS
SOURCE Mycobacterium tuberculosis.
ORGANISM Mycobacterium tuberculosis
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacteriinae; Mycobacteriaceae;
Mycobacterium; Mycobacterium tuberculosis complex.
1 (bases 1 to 474)
Verdon,A., Hartskeerl,R.A., Schultema,A., Kolk,A.H., Young,D.B. and
Lathigra,R.
The 14,000-molecular-weight antigen of Mycobacterium tuberculosis
is related to the alpha-crystallin family of low-molecular-weight
heat shock proteins
J. Bacteriol. 174 (4), 1352-1359 (1992)
92138631
Genbank staff at the National Library of Medicine created this
entry [NCBI g1bbsq 79751] from the original journal article.
This sequence comes from fig2.
FEATURES
    source          1..474
                     Location/Qualifiers
                     /organism="Mycobacterium tuberculosis"
                     /db_xref="taxon:1773"
                     /gene="14 kda antigen"
                     /note="this sequence comes from fig2"
                     /codon_start=1
                     /protein_id="AB21317.1"
                     /db_xref="GI:244562"
                     /translation="MATTLPVQRHRSLEPPESEFAAPPSFAGLRPTEDRLRMLED
EMKEGRYERVALEPVDKDVDMVRGQLTIKARETQKDFDGRSEFAVGSFVRTV
SLPVGADDDDKATYDKGILTVSVAVSGKPTKEKHQIRSTN"
BASE COUNT      94 a 148 c 142 g 90 t
ORIGIN

```


Query Match 99.3%; Score 428.8; DB 73; Length 474;
 Best Local Similarity 99.5%; Pred. No. 5, 4e-68;
 Matches 430; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY	1	gcccacccctccggttcaagcgcacccgcggttccttcccgagtttctgagctg 60
DB	19	gcccacccctccggttcaagcgcacccgcggttccttcccgagtttctgagctg 78
QY	61	ttcgagccttccggttccttcccgagtttctgagctg 120
DB	79	ttcgagccttccggttccttcccgagtttctgagctg 138
QY	121	ctggaagacagatgaagaagagcgctacgagtgacgcgaggttccggggtcgcac 180
DB	139	ctggaagacagatgaagaagagcgctacgagtgacgcgaggttccggggtcgcac 198
QY	181	ccgcgaagagacgtccacattatgtctcgatgtgtcaactgaagagcgagcg 240
DB	199	ccgcgaagagacgtccacattatgtctcgatgtgtcaactgaagagcgagcg 258
QY	241	accgagcaaaagacttaagcgtgctcggaatttcgttaaggttccttcgttcgaag 300
DB	259	accgagcaaaagacttaagcgtgctcggaatttcgttaaggttccttcgttcgaag 318
QY	301	gtctcgtcgcgtgaagtgctcgcagagacatttaagccacccacagagagcatt 360
DB	319	gtctcgtcgcgtgaagtgctcgcagagacatttaagccacccacagagagcatt 378
QY	361	cttactgtgtcgtggttcgcgaagagagccacccgaaagacattcagatccgg 420
DB	379	cttactgtgtcgtggttcgcgaagagagccacccgaaagacattcagatccgg 438
QY	421	tcacccaactga 432
DB	439	tcacccaactga 450

RESULT 3
 MTVO18/c 53450 bp DNA BCT 18-JUN-1998
 LOCUS Mycobacterium tuberculosis H37Rv complete genome; segment 90/162.
 DEFINITION AL021899 AL123456
 ACCESSION AL021899.1 GI:3242282
 VERSION
 KEYWORDS
 ORGANISM
 SOURCE
 Mycobacterium tuberculosis.
 Mycobacterium tuberculosis
 Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 Actinomycetales; Corynebacterineae; Mycobacteriaceae;
 Mycobacterium.
 1 (bases 1 to 53450)
 Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C.,
 Harris, D., Gordon, S.V., Eigemeier, K., Gas, S., Barry III, C.E.,
 Tekle, A.F., Badcock, K., Basham, D., Brown, D., Chillingworth, T.,
 Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S.,
 Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J.,
 Moule, S., Murphy, L., Oliver, S., Osborne, J., Quail, M.A.,
 Rajandream, M.A., Rogers, J., Rutter, S., Seeger, K., Skelton, S.,
 Squares, S., Squires, R., Sulston, J.E., Taylor, K., Whitehead, S. and
 Barrell, B.G.
 TITLE
 Deciphering the biology of Mycobacterium tuberculosis from the
 complete genome sequence
 JOURNAL
 Nature 393 (6685), 537-544 (1998)
 MEDLINE
 9825967
 ERATUM: [[published erratum appears in Nature 1998 Nov
 12; 396(6707):190]]
 REFERENCE
 2 (bases 1 to 53450)
 Parkhill, J.
 Direct Submission
 Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium
 tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome
 Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique
 Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux,
 75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk

COMMENT

On Jun 20, 1998 this sequence version replaced gi:2896750.
 Notes:

Details of M. tuberculosis sequencing at the Sanger Centre are
 available on the World Wide Web.

(URL, http://www.sanger.ac.uk/Projects/M_tuberculosis/) CDS have
 been renumbered from the original cosmid submissions but the old
 gene designations are in brackets after the new gene numbers.
 Gene prediction was based on a Hidden Markov Model of TB genes
 implemented in TParse (Krogh) supplemented with visual inspection
 of positional base preference in codons, especially where there is
 an increase in the observed/expected third position G + C.

CAUTION: In some cases we may not have predicted the correct
 initiation codon. Where possible we choose an initiation codon
 (atg, gtg, or ttg) which is preceded by an upstream ribosome
 binding site sequence (optimally 5-13bp before the initiation
 codon). If this cannot be identified we choose the most upstream
 initiation codon.

FEATURES

source

1..53450
 /organism="Mycobacterium tuberculosis"
 /strain="H37Rv"
 /db_xref="taxon:1773"

misc_feature

<1..51145
 /note="fragment designated v018. Does not represent a
 physical clone"

gene

complement(114..1370)
 /gene="Rv2015c"

CDS

complement(114..1370)
 /note="Rv2015c"

/note="Rv2015c" (MTVO18.02c), len: 418. Nearly identical
 to Mycobacterium tuberculosis Rv1765c (MTVC28.31c, 378 aa),
 anorf starting next to ISB9, and ending in IS6110.
 Different N-terminus Chosen and C-terminus differs as that
 of MTVC28.31c has been truncated by IS6110. Does NOT show
 similarities with transposons. BLAST hits with non-IS part
 of MTU78639. FASTA scores: 2958901/MTVC28.31 (378 aa) opt:
 2417 z-score: 2553.1 E(): 97.8% identity in 364 aa
 overlap. TParse score is 0.939"

/codon_start=1

/transl_table=11

/product="hypothetical protein Rv2015c"

/protein_id="CA17229.1"

/db_xref="GI:2896752"

/db_xref="SPTREMBL:O53461"

/translation="MSTATSGAAYVSPAERYEVLEELAEIAGRNAIDGRIVEIVA
 ELDRDGLMGVYAGSADAHYAVKMGSSGNHTATYARIPPEPRCARMGREGTSL
 DOVGATACRAGSADAHYAVKMGSSGNHTATYARIPPEPRCARMGREGTSL
 DEOFCWRIKILPHVEAAKFDALOSHDALAEYRHDNDGVSODRPPICGNEAF
 LRVVAGMDAIVARPRGHTTYVNHLDYQERAGLHGLPLSESEKRYLLCDATFEA
 WERDGVYIGGRTTRQINRRRLRLERDRFCVPGCGATRGHAAHRIHWDGAT
 ELANLVICPYHRAHNRGLITITGPANRLTVADSDAGPLISAGSLARASTRPDAVAP
 WPGPGERADWWMYEPPQPPPSLN"

1715..1719

/note="possible RBS, ggcga, for Rv2016"

1724..2299

/gene="Rv2016"

1724..2299

/gene="Rv2016"

1724..2299

/note="Rv2016"

1724..2299

/note="Rv2016"

1724..2299

/note="Rv2016"

1724..2299

/note="Rv2016"

1724..2299

/note="Rv2016"

1724..2299

/note="Rv2016"

1724..2299

/note="Rv2016"

1724..2299

/note="Rv2016"

1724..2299

/note="Rv2016"

1724..2299

/note="Rv2016"

1724..2299

/note="Rv2016"

1724..2299

/note="Rv2016"

gene 2296. .3336
/gene="Rv2017"
CDS 2296. .3336
/note="Rv2017", (MTV018.04), len: 346. Unknown but shows
similarity at N-terminal end to several transcriptional
regulators e.g. Bacillus subtilis BSUB0012.44 (108 aa).
ContainsSP00142 Neutral zinc metalloproteinases,
zinc-binding regions in C-terminal half, may be
fortituous. FASTA scores: 299115BSUB0012.44 Bacillus
subtilis (108 aa) opt: 154 z-score: 199.9 E(): 0.0012;
35.5% identity in 62 aa overlap. Tbpase score is 0.908.
Contains probable helix-turn-helix motif at aa 18-39
(Score 2243, +6.83 SD)
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv2017"
/protein_id="CA117231.1"
/db_xref="GI:2896754"
/db_xref="SPTREMBL:O53463"
/translation="MNGLDGLVAVARRKAGLTQIELAEVLGTPAIPNRYSGDRDPD
OHTVAKTLEIGTDDLLHGNFRGALAVADAMRRHKTKKASAMOLEARLNILRVH
ASFLREVAINSQHYPADEPTAEDARLYRAQRMPKGVVULTRMEAGCLV
FEEDFATQRIIDGLSOWDDYVWMLINANAPDKRLTLAEHLVHSTNPENET
EATFAAEFLMPESEIRPELRDLGLLELKEKWCVSQALLARVRLGLVSAEART
KLYKANAGWKTKPEGISIVREKPSLPHIGIMTLRSRGFTDQAAALAGYANPADN
PEREGGRLHAI"
2884. .2913
/gene="Rv2017"
/note="P500142 Neutral zinc metalloproteinases,
zinc-binding region signature"
3565. .3569
/note="possible RBS, 99ag9, for Rv2018"
3578. .4297
/gene="Rv2018"
3578. .4297
/note="Rv2018"
/note="Rv2018", (MTV018.05), len: 239. Unknown but similar
to MTCY339.01c (238 aa). FASTA scores: 277163IMTCY339.1
Mycobacterium tuberculosis cosmid (238 aa) opt: 142
z-score: 175.1 E(): 0.029; 24.8% identity in 250 aa
overlap. Tbpase score is 0.929. Contains probable
helix-turn-helix motif at aa 215-236 (Score 1175, +3.19
SD)
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv2018"
/protein_id="CA117232.1"
/db_xref="GI:2896755"
/db_xref="SPTREMBL:O53464"
/translation="MAGDQELERFDVPLTYLAESRYLVPRATLATWADGERRPA
NAPVAGOPITIALPPTGSHARLPVGLAEAVLAPFRAGVPMKRIPELDLTKN
VGFHALASODLCTDGAELVLMRFAERSGSDPLVYRGLIVPSGQYVKEIYEHLTQ
QISFADNPLASMKRLPQYGDANVYLDPRRGQPVDSGVRADVLGLRAGATITQA
VADDIVTPDQLDALDAITAA"
4287. .4703
/gene="Rv2019"
4287. .4703
/note="Rv2019", (MTV018.06), len: 138. Unknown"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv2019"
/protein_id="CA117233.1"
/db_xref="GI:2896756"
/db_xref="SPTREMBL:O53465"
/translation="MOPDRNLADLDHITVDSLGAQVQPOLRIDAGFLLTTRREHYG
ETQAOVSQDHKMIAMTAEQGWGFHNDANIIRNVAERVRVLDGTGARLFCVPRADIIAE
QVAAIRIASLIAIARAARPPGFIYVHPSKIVRYL"
complement(4719. .5018)
/gene="Rv2020c"
complement(4719. .5018)
/gene="Rv2020c"

Query Match 99.3%; Score 428.8; DB 2; Length 53450;
Best Local Similarity 99.5%; Pred. No. 1,2e-68;
Matches 430; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 gccacacacccctccggttcagcgccacccggttcctctcccgagttctcgaacg 60
17227 GCCACCAACCTTCCCGTTTCAGCGCCACCCGCGCTTCTCCCGAGTTTCTGACCTG 17168
Qy 61 ttcgagcctcccgatcgcgcgagtcgcgcacaccccttcgaacccggttcgagc 120
17167 TTCGCGCCTTCCCGTTCATTCGCGGACTCCGCGCCACTTTCACACCCGTTGATCGG 17108
Qy 121 ctggaagcagagatgaagaggcggttcgaggttcgcgcgcgcgcgcgcgcgcgcgc 180
17107 CTGGAACACAGATGAAGAGGCGGCTTACAGGTACGCGGAGCTCCCGGAGTGCAC 17048
Qy 181 ccgacaagagcgtccacatcgttcggttcgagtcgagtcgacatcgaagcgagcgc 240
17047 CCGACACAGAGCGTCGACATTTAGTTCGCGGATGCTGACGTGACATCAAGCGGACCGC 16988
Qy 241 accgagcagaagactaagcgcgttcgcgaatcgcgcgagtccttcgttcgacg 300
16987 ACCGAGCAGAAGACTTTCGACGCGTCCGCAATTCGCGTTCGTTCCGTCGACG 16928
Qy 301 gttcgctcgcggttaggttcgaggaagacattaaagccacacgaagggcatt 360
16927 GTTCGCTCGCGGTAGGTCTGACGAGACGACATTAAAGCCACCTACGACAAAGGCAAT 16868
Qy 361 ctactctgcggttcgaggttcggaaggaacccaacgaagacatcagatccg 420
16867 CTTACTGTGCGGTGCGGCTTCGGAAGGACCAACCAAGCAATTCAGATCCGG 16808
Qy 421 tccacgaactga 432
16807 TCCACCAACTGA 16796
Db 16807 TCCACCAACTGA 16796
RESULT 4
AF072875 528 bp DNA BCT 16-JUN-1999
LOCUS Mycobacterium smegmatis putative HSP20 (hsp) gene, complete cds.
DEFINITION AF072875
ACCESSION AF072875
VERSION AF072875.1 GI:5070145
KEYWORDS
SOURCE Mycobacterium smegmatis.
ORGANISM Mycobacterium smegmatis
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacterineae; Mycobacteriaceae;
Mycobacterium.
REFERENCE 1 (bases 1 to 528)
Caceres,N.E., Feng,Z., Li,L.-L., Kapur,V., Cirillo,J.D. and
Barietta,R.G.

TITLE Identification of a HSP20 heat shock protein homolog from
JOURNAL Mycobacterium smegmatis
REFERENCE Unpublished
AUTHORS 2 (bases 1 to 528)
Caceres, N.E., Feng, Z., Li, L.-L., Kapur, V., Cirillo, J.D. and
Barletta, R.G.
TITLE Direct Submission
JOURNAL Submitted (18-JUN-1998) Veterinary and Biomedical Sciences,
University of Nebraska-Lincoln, Rm 211 VBS, Fair Street and East
Campus Loop, Lincoln, NE 68583-0905, USA
FEATURES
source
1. .528
/organism="Mycobacterium smegmatis"
/strain="mc(2)155"
/db_xref="taxon:11772"
53. .502
/gene="hsp"
53. .57
/gene="hsp"
65. .502
/gene="hsp"
/note="heat shock protein"
/codon_start=1
/transl_table=11
/evidence-not_experimental
/product="putative HSP20"
/protein_id="AND39038.1"
/db_xref="GI:5070146"
/translation="MSTLMKTPAVVTRPAMOLDNNVRDFPGPADMFKGFPAEYTR
DGEDAVSELPVGVGKDVNVEVGRGLVHGERDERSESDGRTYSEVAYGSFRR
BEKLPAAHTGDAVTASYDAGVLYKRVAGVAGEOPRIARESR"
BASE COUNT 103 a 167 c 184 g 74 t
ORIGIN
Query Match 11.2%; Score 48.2; DB 1; Length 528;
Best Local Similarity 51.6%; Pred. No. 16;
Matches 144; Conservative 0; Mismatches 123; Indels 12; Gaps 1;
QY 124 gaagacgagatgaagaaggcgctacgaggtacgacgagagcttccggggtcgacccc 183
DB 185 GAGGTCTCAAGAGACGCTGAGGACGCCGTGTGTCGCGAAGCTCCCGTGGATGTC 244
QY 184 gacaaagacgtcacatltatgtcgcgcatgtgtcagctacacatcaagccgag----- 227
DB 245 GCGAAGACGTCAACCTCGACGAGGTGACCGCGCGGTGTGTCATCCACGCGGCGACGCGT 304
QY 238 -----cgaccgagacgaagacttagacggtcgtcgtgaattcgcgtacggttccttc 291
DB 305 GACGAACGCTCCGAGGAATCCGACGCGCGCAGCTACAGTAGAGTGCCGCTACGGCTCTTC 364
QY 292 gtctgacgagtgctgcgtcgcggtagtgctgacgagacgaacatgaagccactacgac 351
DB 365 CCGCGGAGCTTCAAGCTGCGCCGCGACGCTCACGGGCGGCGGCTACCGCTCTACGAC 424
QY 352 aagggactcttactgtgtcgtgtgaggttcggaagg 390
DB 425 GCGAGCGCTACTGAGGTACGGGTGCGCGGTGCTTACGCG 463

REFERENCE 1 (bases 1 to 2982)
AUTHORS Zhang, W., Jiang, W., Zhao, G., Yang, Y. and Chiao, J.
TITLE Sequence analysis and expression of the aspartokinase and aspartate
semialdehyde dehydrogenase operon from rifamycin SV-producing
amycolatopsis mediterranei
JOURNAL Gene 237 (2), 413-419 (1999)
MEDLINE 99453302
REFERENCE 2 (bases 1 to 2982)
AUTHORS Zhang, W., Jiang, W.H., Zhao, G.P., Yang, Y.L. and Chiao, J.S.
TITLE Direct Submission
JOURNAL Submitted (13-MAR-1999) Microbiology, Shanghai Institute of Plant
Physiology, Academia Sinica, 300 Fenglin Road, Shanghai, SH 200032,
P.R.China
FEATURES
source
1. .2982
/organism="Amycolatopsis mediterranei"
/strain="Uj2"
/db_xref="taxon:33910"
220. .2525
/gene="aspartokinase and aspartate-semialdehyde
dehydrogenase operon"
220. .1485
/gene="aska"
220. .1485
/gene="aska"
/codon_start=1
/transl_table=11
/product="aspartokinase subunit A"
/protein_id="AAD49567.1"
/db_xref="GI:5733411"
/translation="MAIVQKYGSSLSADRIKRYARIVATTKKAGNDVYVVCAMG
DTTDELIDAOQVNPAPPREKMDLITIGERISNSLVAAIAAGAEASFSGAGV
VTTSVGNARITIDVTSRVTEALDQGYIALVAGFQGVADDTITTLGSGSDTTAVA
IAAALNADVCIELYSDVDGYTADPRVPDAKRLDVTYEEMLELAAGSKILHLSVE
YARRYGVPPIRVSSYSDKPGTGTGSIETIPVEQALITYGAHDSREAKITVGPDDHT
GAARLFRVITADAEIDIMVLONVSTVSGRFDITFTLSKAGAAVARELEKVOAIG
FESVLYDHDVGVGVYVAGMRSHPTATFCCALAEAGVNIETINTSEIRISVILRDA
QDDAVRAIHEHFEELGDEAVVAGSGR"
967. .1485
/gene="aska"
/note="orf2"
/codon_start=1
/transl_table=11
/product="aspartokinase subunit B"
/protein_id="AAD49568.1"
/db_xref="GI:5733412"
/translation="MEQALITGVADHRSKATITVGPDDHTGAARIFPVIADEIDI
DMVLONVSTVSGRFDITFTLSKAGAAVARELEKVOAIGFESVLYDHDVGVYVAG
ACMRSHPTATFCCALAEAGVNIETINTSEIRISVILRDAQDDAVRAIHEHFEELG
DEAVVAGSGR"
1485. .2525
/note="orf3"
/codon_start=1
/transl_table=11
/product="aspartate-semialdehyde dehydrogenase"
/protein_id="AAD49569.1"
/db_xref="GI:5733413"
/translation="MADGIRVGVATGCVGAVMRKLLAEREFPYAEIRYFASARSAG
SKLPWRDEITIEDASTDPSGLDIALFSAGSGTCAQAPFAAGAVYVINDSSAFRM
DDVDPLVWSEVNEAVKEARKGILIANPQCTTIAAPVAVLKPLHDEAGLVRLVASTQVA
SSGGLAGVDELAGVRAAEHSLTHGCAATDPKPKERYVPIAFNVLPAGSIVYD
GFEFTDEKRRNRSKRLTSLTGLAVSCTCAVVPVFSGSVSNVEFEKPLSVVERATE
LITHAAGVELSEEPPIQAAAGNDPSYVGRIRVPDCEGGRGIALFLSDNLRKGAALN
AIQIAELVAQOI"
BASE COUNT 472 a 996 c 1072 g 442 t
ORIGIN
Query Match 11.1%; Score 47.8; DB 1; Length 2982;
Best Local Similarity 48.4%; Pred. No. 11;
Matches 164; Conservative 0; Mismatches 172; Indels 3; Gaps 1;
QY 51 ttctgagctgttcgcgcttccgcgtatcgcgcgacccgacacttgcacacccg 110

Db 510 TTGCGAGGCGGCGTCTCAGACGTGGTGCACGAGCGGCGGCGATCGACGTAC 569
 111 GTTGATGCGGCTGGAAGAGATGAAGAGGCGGCTGAGGAGGCGGAGGCTTC 170
 Db 570 GCGGAGCGGCGGTACCGAGGCGCTCGACGAGGGGTACATGCGGCTGCGGCTTCCA 629
 QY 171 cggagtcgacccga---caagagcgtccacatgatgtccgagatgtacgtaccat 227
 Db 630 GGGCGTGGCGGACGACCAAGGACATACCGCTGGCGGCGGCGGCTCGGACACAC 689
 QY 228 caagcccgagcgcacgcagagacgaagactagagtcgctcggaatcgctagcttc 287
 Db 690 CGCGTGGCGCTGCGGCGGCGCTGACGCGGCTGCGGAGATCTCCGATGTGA 749
 QY 288 ctccgttcgacggtgtcgttcgctgaggtgtcgaagagacgaactaagccacta 347
 Db 750 CGGTGTGACAGCGGCGGCGGCGGCTGCGGCGGCGGCGGCGGCTGCGGCTGCGG 809
 348 cgacaagggacatcttactgtgtcgtgtcgttcgga 386
 810 GTACGAAGAGATGCTGAGCTCGCGGCGGCGGCTGCAA 848

RESULT 6
 LOCUS MTU034 22800 bp DNA BCT 18-JUN-1998
 DEFINITION Mycobacterium tuberculosis H37Rv complete genome; segment 13/162.
 ACCESSION AL021929 AL123456
 VERSION AL021929.1 GI:3242291
 KEYWORDS
 SOURCE
 ORGANISM

Mycobacterium tuberculosis.
 Mycobacterium tuberculosis
 Bacteria: Firmicutes: Actinobacteria: Actinobacteridae:
 Actinomycetales: Corynebacterineae, Mycobacteriaceae;
 Mycobacterium.

REFERENCE 1 (bases 1 to 22800)
 AUTHORS Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C.,
 Harris, D., Gordon, S.V., Eigemeier, K., Gas, S., Barry III, C.E.,
 Tekela, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T.,
 Connor, R., Davies, R., Devlin, K., Fellwell, T., Gentles, S.,
 Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J.,
 Moule, S., Murphy, L., Oliver, S., Osborne, J., Quail, M.A.,
 Rajandream, M.A., Rogers, J., Rutter, S., Seeger, K., Skelton, S.,
 Squares, S., Squires, R., Sulston, J.E., Taylor, K., Whitehead, S., and
 Barrell, B.G.
 Deciphering the biology of Mycobacterium tuberculosis from the
 complete genome sequence
 Nature 393 (6685), 537-544 (1998)
 98295987
 Erratum: [[published erratum appears in Nature 1998 Nov
 12;396(6707):190]]
 2 (bases 1 to 22800)
 PARKHILL, J.

REFERENCE 2
 AUTHORS Parkhill, J.
 TITLE Direct Submission
 JOURNAL Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium
 tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome
 Trust Genome Campus, Hinxton, Cambridge CB10 1SA unite de Genetique
 Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux,
 75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk
 On Jun 20, 1998 this sequence version replaced gi:2909438.
 Notes:
 Details of M. tuberculosis sequencing at the Sanger Centre are
 available on the World Wide Web
 (URL: <http://www.sanger.ac.uk/Projects/M.tuberculosis/>) CDS have
 been renumbered from the original cosmid submissions but the old
 gene designations are in brackets after the new gene numbers.
 gene prediction was based on a Hidden Markov Model of TB genes
 implemented in TParse (Krogh) supplemented with visual inspection
 of positional base preference in codons, especially where there is
 an increase in the observed/expected third position G + C.
 CAUTION: In some cases we may not have predicted the correct
 initiation codon. Where possible we choose an initiation codon

FEATURES
 source (atg, gtg, or ttg) which is preceded by an upstream ribosome
 binding site sequence (optimally 5-13bp before the initiation
 codon). If this cannot be identified we choose the most upstream
 initiation codon.
 Location/Qualifiers
 1..22800
 /organism="Mycobacterium tuberculosis"
 /strain="H37Rv"
 /db_xref="taxon:1773"
 <1..22238
 /note="fragment designated v034. Does not represent a
 physical clone"
 184..1350
 /gene="lpgI"
 184..1350
 /note="lpgI"
 /note="RV0237, (MTU034.03), len: 388. Lipoprotein similar
 to ALTEROMONAS SP. BETA-HEXOSAMINIDASE A PRECURSOR (598
 aa) and SYNECHOCYSTIS SP. PCC6803 sptr|P74340|P74340
 BETA-GLUCOSIDASE (538 aa). Contains signal sequence and
 appropriately positioned PS00013 prokaryotic membrane
 lipoprotein lipid attachment site. FASTA scores
 sp|P46823|HEXA_ALTRSO BETA-HEXOSAMINIDASE A PRECURSOR (598
 aa) opt: 415 z-score: 439.1 E(): 5.6e-17; 31.28
 identity in 343 aa overlap; and sptr|P74340|P74340
 BETA-GLUCOSIDASE (538 aa) opt: 414 z-score: 438.7 E():
 6.1e-17; 30.6% identity in 320 aa overlap. TParse score
 is 0.890"

misc_feature

gene

CDS

misc_feature

RBS

gene

CDS

gene

CDS

/codon_start=1
 /transl_table=11
 /product="lpgI"
 /protein_id="CAI17329.1"
 /db_xref="GI:2909441"
 /db_xref="SPRMBL:053660"
 /translation="MAFPPTLALIAAALVYACSGGTPGSSITSGASPTAPVAV
 VPSKAEPAIGIPALSLPRDLQALVGVADADAAVAVTVNHVGIILGSDITLTF
 DGLAEIVAGGGLPLAVSDEGGRVSRISLIGTGSAARELAQVOYRDITF
 DRROMRKLGITIDFAPVVDVDAPODTYIGDFSFPDPAIVAGAVAGCIRLDGV
 LPLKHPGHRGSGSHNGVYTPPLDLVGDVLPYRTLVQAVGVVGHVLQVGV
 LNSSEPASLSKAANVNLRTGTGAPPPGPPVSDLSGMAASDFGSENVLRILQ
 AGDILAWVTTEVPALVDRLEQALRAGELPMSAVDRSVRAATMGPPGCR"
 211..243
 /gene="lpgI"
 /note="PS00013 prokaryotic membrane lipoprotein lipid
 attachment site"
 1414..1419
 /note="possible RBS, aagggg, for RV0238"
 1426..2040
 /gene="RV0238"
 1426..2040
 /gene="RV0238"
 /note="RV0238, (MTU034.04), len: 204. Possible regulatory
 protein, similar to Mycobacterium tuberculosis
 MTCY07D11.18c(228 aa) and MTCY10G2.30c (197 aa). FASTA
 scores 295120(MTCY7D11_18 (228 aa) opt: 266 z-score:
 346.6 E(): 8.3e-12; 28.1% identity in 196 aa overlap.
 TParse score is 0.876"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein RV0238"
 /protein_id="CAI17330.1"
 /db_xref="GI:2909442"
 /db_xref="SPRMBL:053661"
 /translation="YAGGTRRLPRPRAREQMDAAYQMSVNCYHETSMDAIAEAOI
 SKPMILVYGSKEDELFGACLNPRMSFFIDALNSSINFDOSPDLNNTVATIAEAOI
 NRSWLVMTQTQATISQAFATVAREGQIVQVLAELVARGTRGRLDADVIEMAAVLA
 GGEAVATRLIGDITDVDEAAEMIMLFWLGLKAPVDRLELTHG"
 2102..2335
 /gene="RV0239"
 2102..2335
 /gene="RV0239"
 /note="RV0239, (MTU034.05), len: 77. Unknown but similar
 to Mycobacterium tuberculosis protein MTCY359.34 (87 aa).


```

RBS      475..478
         /gene="sigma"
         /note="putative"
gene      475..2657
         /gene="sigma"
         /gene="sigma"
         /gene="sigma"
         /standard_name="sigma"
         /codon_start=1
         /transl_table=11
         /product="sigma factor"
         /protein_id="AAA26473.1"
         /db_xref="GI:152656"
         /translation="MPTQPSKSGSKPKKKVDPYIRKKKAPDSYAKTKAAAAEEL
         ASAEATEKALELKKRKRAVTEASADDDPEEAAEAAAEVDPAAVEDEEETIAE
         RKWEKDLAAGEREGFLTYDEVNDALPDADIVSDQIDVMSMFGNDIEIYDAOKAAQ
         SNEIKPTVAVEESKODADDDDEDEDEGKSNPVRLYLRKMGSVLLRREGVEEI
         AKRIEDGEVLRAMACRAVADILDIGNKTGKLRVDYIKDAPEEAAEGAEAA
         PEEGVEGPOOLAOSELNKIEQINKOIERFRKFAKDCLEVEELSGKKKLLDVYKKE
         VKQEKDLRTKMEVLEEMRLNKKQVDYIVYINIKOLIERVYKAEELDLRRNGVSM
         KELRPOLKESRDNPYIAKLOKQLENTPEQLDALDRDYTAVRKIKRYEENLIPVDA
         LRRNDARILEERRAERAKSELVEANRLVSIKAKYNNRGLQFLDLIQEENIGLMA
         VDRKPKYKGFSTYATWMIROAITRAIADQARTIRIPHMIEIINKLIRTSRYLVE
         IGRPTPEIEIKEMELPLDKYRKVLKIAKEPISLETPIGEEDSHLGDPIEDKSLVP
         ADAYINMMLARTKRKVLATLPREKVKLRMFRGIGEKSDHTLLEVGDFEYTERIRQ
         IEAKALRKLRRPSSKRLRFVES"
terminator 2635..2657
         /gene="sigma"
         /standard_name="rho-independent"
         /note="putative"
BASE COUNT 601 a 860 c 936 g 351 t
ORIGIN

Query Match      10.0%; Score 43.2; DB 73; Length 2748;
Best Local Similarity 51.6%; Pred. No. 74;
Matches 99; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

QY 64 gggccctccgctcattcgcggaactcgggcccaaccttgacaaccccggtgagcgctg 123
    ||||| || || || || || || || || || || || || || || || || || || ||
DB 1200 GGGCGCGTGGCGACATCCTCGACATCGGCACACAGCTCAAGCGGCGAAGCGCGCGT 1259
    ||||| || || || || || || || || || || || || || || || || || || ||
QY 124 gaagacgagataaagaggagcgctacgaggtacgagcgagcgtcccgagggttcgaccc 183
    ||||| || || || || || || || || || || || || || || || || || || ||
DB 1260 CCCGAGGTATCAAGACCGCGCCGGAAGGCTCAAGCGCGGAGGCGCCGAGAGGCGCC 1319
    ||||| || || || || || || || || || || || || || || || || || || ||
QY 184 gacaaggacgtccacattatgtccgcgattgctcagctgacatcaagcgagcgaccc 243
    || || || || || || || || || || || || || || || || || || || || ||
    1320 GAGGAGGCGGCTCGAAGGGAGCGCCCGACAGCTGGCCGACAGGCGAGCTCAACAAGATC 1379
QY 244 gaggcgaaggac 255
    ||||| || || || || || || || || || || || || || || || || || || ||
DB 1380 GAGCAGATCAAC 1391

RESULT 12
AF257177 1750 bp DNA BCT 09-MAY-2000
LOCUS Streptomyces thermocyanoeoviolaceus xylose isomerase (xyIA) gene,
DEFINITION complete cds.
ACCESSION AF257177
VERSION AF257177.1 GI:7739711
KEYWORDS Streptomyces thermocyanoeoviolaceus.
SOURCE Streptomyces thermocyanoeoviolaceus.
ORGANISM Streptomyces thermocyanoeoviolaceus
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomyicinae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 1750)
AUTHORS Kwak,Y.-Y., Heo,G.-Y., Shin,J.-H., Jang,H.-S., Joo,G.-J. and
Rhee,I.-K.
TITLE Streptomyces thermocyanoeoviolaceus xyIA (xylose isomerase)
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1750)

```

```

AUTHORS Kwak,Y.-Y., Heo,G.-Y., Shin,J.-H., Jang,H.-S., Joo,G.-J. and
Rhee,I.-K.
TITLE Direct Submission
JOURNAL Submitted (17-APR-2000) Department of Agricultural Chemistry,
College of Agriculture, Kyungpook National University, #1370
Sankyuk-dong, Buk-ku, Taegu 702 - 701, Korea
FEATURES
source
    1..1750
        /organism="Streptomyces thermocyanoeoviolaceus"
        /strain="KCCM 40049"
        /db_xref="taxon:106355"
        424..1590
        /gene="xyIA"
        424..1590
        /gene="xyIA"
        /codon_start=1
        /transl_table=11
        /product="xylose isomerase"
        /protein_id="AAF68977.1"
        /db_xref="GI:7739712"
        /translation="MNYQPTPEDRFTGLTWGQGRDPFGDARIPALDPVAVQRLA
        ELGAVGVFHDLDLIPGSSDAEREAHYKRRQALDAGMTVPMAITNLFTII-VFKDG
        AFTANDRVRRTALRKTIKRLNDLAVELGAKYVAMGREGAESGAAKVDRAALDRKE
        AFDLGEVTSQGVDIRAIEPKRPEPRGDLITPTIGHALAFIELEPEELGYNPEV
        GHEOMAGINPFGHIGQALMAKRLPHIDNGSGCIKYODLIRFGADLRAPALVDLLE
        SAGWGPHEPFKPRPTEDIDGWAASAGCRRNLTILKERAARADPEVDALRAAR
        LDOLAEPYAADGLQALLADRYAFEDFEDVDAARAAGMAFERLDOLAMHLLGARG"
BASE COUNT 247 a 661 c 588 g 254 t
ORIGIN

Query Match      9.9%; Score 42.8; DB 1; Length 1750;
Best Local Similarity 57.5%; Pred. No. 1e+02;
Matches 77; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

QY 151 gggtagcgcgagcgttcgccggggtcgaccccgcaagcgtccacattatgttcgcg 210
    ||||| || || || || || || || || || || || || || || || || || || ||
DB 1396 GAGGTCCAGAGAGGCGCTGCTGCTGCCGCGCTCGACACGCTGCCGAGCCGCGCGGC 1455
    ||||| || || || || || || || || || || || || || || || || || || ||
QY 211 gatgttcagctgacatcaagcgagcgacccgacgacgaaggacttagacgtcgtcg 270
    || || || || || || || || || || || || || || || || || || || || ||
DB 1456 GACGGCGCTCAGGCCCTGCTGGCCGACCGCACCGCTACGAGACTTCGACGTGACGCG 1515
    || || || || || || || || || || || || || || || || || || || || ||
QY 271 gaattcggtacgg 284
    || || || || || || || || || || || || || || || || || || || || ||
DB 1516 GCCGCGCGCGCGG 1529

RESULT 13
SHU27483 399 bp mRNA VRL 22-NOV-1996
LOCUS Suid herpesvirus 1 glycoprotein gII mRNA, partial cds.
DEFINITION Submitted (23-MAY-1995) Tin-yun Ho, Department of Veterinary
ACCESSION U27483 Medicine, National Chung-Hsing University, 250 Kuo-Kwang Road,
VERSION U27483.1 GI:1304427 Taichung 40227, Taiwan
KEYWORDS Pseudorabies virus.
SOURCE Pseudorabies virus.
ORGANISM Pseudorabies virus.
Virus; dsDNA viruses, no RNA stage; Herpesviridae;
Alphaherpesvirinae; Varicellovirus.
REFERENCE 1 (bases 1 to 399)
AUTHORS Ho,T., Hsiang,C. and Chang,T.
TITLE Characteristics of early transcripts of pseudorabies virus
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 399)
AUTHORS Ho,T.
TITLE Direct Submission
JOURNAL Submitted (23-MAY-1995) Tin-yun Ho, Department of Veterinary
REFERENCE 1 (bases 1 to 399) Medicine, National Chung-Hsing University, 250 Kuo-Kwang Road,
JOURNAL Taichung 40227, Taiwan
FEATURES
source
    1..399
        /organism="Pseudorabies virus"
        /strain="TNL"

```

```

CDS
<db_xref="taxon:10345"
<1..>399
/codon_start=1
/evidence=experimental
/product="glycoprotein gII"
/protein_id="AAB6946.1"
/db_xref="GI:1335775"
/translation="PFGAIAIGLIVLGAIAFLAVYHISLRNPKALYPYTKAL
KEDVESGDDEAKLDQARDIRMTSYSLAEQDEHRRARTAGPSRWAGSGNMLLA
AGTTSASRARTPTCSPLPRTIMRLFGNT"
polyA_signal
372..377
/evidence=experimental
399
polyA_site
/evidence=experimental
71 a 141 c 135 g 52 t
BASE COUNT
ORIGIN
Query Match 9.8%; Score 42.2; DB 72; Length 399;
Best Local Similarity 49.8%; Pred. No. 2.1e+02;
Matches 107; Conservative 0; Mismatches 108; Indels 0; Gaps 0;
Oy 41 tccccgaattctctagactgtctgcggccctccgcgtacatccgcgagctccgcccacct 100
26 Tccttgctgctgcggccgctgctgcggcccttctctgacctaccggacatctgcggccctgcg 85
Oy 101 tcgacaccggtltgatlgcgcttggaagacgaatgaaagagggcgctacgaagtacgcg 160
86 GCCGCACCCCATGAAGGCCCTGTACCCCGTCACAGCAAGAGCGCTCAAGAGGAGGACGGCG 145
Oy 161 cggagctcccggtgtgaccccgacaagagagctccacatattgtctcgatgtgcacg 220
146 TCGAAGAAAGCGACATGACGACGAGCCCAAGCACTGGACACGCCCGGACATGATCCGTTACA 205
Oy 221 tgaccatcaagcgagcgacgcacacggagagaagac 255
Db 206 TGTCCATCTGTGCGGCCCTTCAGACGACGAGACAC 240

RESULT 14
AF257079
LOCUS AF257079 2724 bp DNA VRL 09-MAY-2000
DEFINITION Pseudorabies virus glycoprotein B gII (gB) gene, partial cds.
ACCESSION AF257079
VERSION AF257079.1 GI:7739722
KEYWORDS
SOURCE
ORGANISM
Pseudorabies virus.
Pseudorabies virus.
VIRUSES; dsDNA viruses, no RNA stage: Herpesviridae;
Alphaherpesvirinae; Varicellovirus.
REFERENCE
1 (bases 1 to 2724)
Hong W., Chen H., Xiao S., He Q. and Fang L.
Clone and sequence analysis of gB gene of Pseudorabies Virus Ea
strain
JOURNAL Unpublished
2 (bases 1 to 2724)
Hong W., Chen H., Xiao S., He Q. and Fang L.
Direct Submission
TITLE Submitted (17-APR-2000) College of Animal Science & Veterinary
JOURNAL Medicine, Huazhong Agricultural University, Lion Mountain Street,
Wuhan, Hubei 430070, P.R.China
Location/Qualifiers
1..2724
/organism="Pseudorabies virus"
/strain="Ea"
/db_xref="taxon:10345"
7..>2724
/gene="gB"
7..>2724
/gene="gB"
/codon_start=1
/product="glycoprotein B gII"
/protein_id="AAB6946.1"
FEATURES
source
gene
CDS

```

	Query Match	9.8%; Score 42.2; DB 71; Length 2724;	
	Best Local Similarity	49.8%; Pred. No. 1.1e+02;	
	Matches 107; Conservative	0; Mismatches 108; Indels	Gaps 0;
OY	41 tcccgaagtttctgagctgttcgcgcgtccgccgtcattcgccgactccggccacact	100	
Db	2423 TGCNTGTCCTGGCCGCCGTGGTGGCGGCTTCCTTGCCACCGGCACATCTCGCCCTGCG	2482	
OY	101 tcgaaccgcglttgatgctgcgtctggaagacgagaagaggggcgctacagagtacgcg	160	
Db	2483 GCCCGAACCCCATGTAAGGCGCCTGTACCCTGGTCACGACGAAGCGCTCAAGAAGACGCGCG	2542	
OY	161 cgaagctcccgagggtcgaccaccgacaaggaagcgtccaatcatatggtccgcgatgtagcgc	220	
Db	2543 TCGAAGAGGAGCAGACTGGACGGCCAGGCGCAAGCTGTGAGCACGCCCCGGACATGATCCGGTACA	2602	
OY	221 tgaccataaagccgagcgcgacacgagacgaagagac	235	
Db	2603 TGTCATCGTGTGGCGCCTCGACGACGAGAGAC	2637	
RESULT 15			
SRY15759			
LOCUS	SRY15759	3354 bp	DNA BC ^T 21-AUG-1998
DEFINITION	Streptomyces rochei silent ABC transporter genes.		
ACCESSION	U15759		
VERSION	V15759.1 GI:3451332		
KEYWORDS	ABC transporter; ATPase.		
SOURCE	Streptomyces rochei.		
ORGANISM	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Streptomycetes rochei		
REFERENCE	Actinomycetales; Streptomycinidae; Streptomycetaceae; Streptomyces.		
AUTHORS	1 (bases 1 to 3354)		
TITLE	Fernandez-Moreno,M.A., Carbo,L., Cuesta,T., Vallin,C. and Malpartida,F.		
JOURNAL	A silent ABC transporter isolated from Streptomyces rochei F20 induces multidrug resistance		
MEDLINE	J. Bacteriol. 180 (16), 4017-4023 (1998)		
REFERENCE	9836186		
AUTHORS	2 (bases 1 to 3354)		
TITLE	Malpartida,F.		
JOURNAL	Direct Submission		
FEATURES	Submitted (02-DEC-1997) F. Malpartida, Centro Nacional de Biotecnologia, Campus de la Universidad Autonoma, 28049-Canto Blanco, Madrid, SPAIN		
source	Location/Qualifiers		
	1..3354		
	/organism="Streptomyces rochei"		
	/strain="F20"		
	/db_xref="taxon:1928"		
	1..386		
	/gene="ORFA"		

```
CDS
<1..388
/gene="ORF1"
/codon_start=2
/transl_table=11
/protein_id="CAA75760.1"
/db_xref="GI:3451323"
/db_xref="SPTREMBL:O86915"
/translation="SSAMATHAAVPHRHGALGVVYPLTIGALVGLTSPFMTRENGAS
AGAOFAVCIVSAVMAGLCYAMGRFOHMPREVRAAATVAVLMGCAVGLTSLSGNSTL
TIVVLSLAIAAATLVITFYIYTHED"
Join(548..554,561..1652)
/gene="ORF1"
548..554
/gene="ORF1"
561..1652
/gene="ORF1"
/function="traffic ATPase"
/codon_start=1
/transl_table=11
/protein_id="CAA75761.1"
/db_xref="GI:3451324"
/db_xref="SPTREMBL:O86916"
/translation="WSTELAIETGVLVFGSENAVDGVDLRVPAGTYVGLGPNAG
KTTYKVATLRLRDPGEARVGRDYADADAVRVSALTQYASVDRLTGTENLL
LARLGHSKPAAREAAQLDAFGLSEAGROVKSYSGMRBRIDIASILNTPLLF
LDEPTGLDPSRNOVDIVRAVAGCTVLTTOYIDEADOLASRIAVTDHGRYIAE
GTFGLKASVSGAVHLRLREAEORPEERILALANTVOILDADVALTRVNGHGT
EDGAENARALSELARAGITVDNFALQPSLDEVELALTOKKALNGSDADMDNGANGA
NGANGSANGNSANGMTAEOKETAV"
1633..1641
/gene="ORF2"
1633..1641
/gene="ORF2"
1649..2503
/gene="ORF2"
1649..2503
/gene="ORF2"
/function="transmembrane protein"
/codon_start=1
/transl_table=11
/protein_id="CAA75762.1"
/db_xref="GI:3451325"
/db_xref="SPTREMBL:O86917"
/translation="MSTVATKEDHSTFAPRAEDLALILVAGDPLRPSAISITF
GWRMLKIKHYPEOLFDTYARPVMMVLMYTLFGCALGSDATYIOELLPGILMSV
MITMTGVAINTDIDKGVDRFRLPIWRPAPVGYLLGDVRYFIVASVMLTGVVI
GYRGGATGILGLALMLFSFSWIMTGLMLRTEKSMVSMNVIFPLTFLSN
VFVDPRTMPGWLQAFVNNSPYTHLATAVREIMAGNMPADIAMSLIGMSARFVLVGAV
TMRLYNRK"
complement(2580..2816)
/gene="ORF3"
complement(2580..2789)
/gene="ORF3"
/codon_start=1
/transl_table=11
/protein_id="CAA75763.1"
/db_xref="GI:3451326"
/db_xref="SPTREMBL:O86918"
/translation="MRIGSGNNAYTHIIVAIGDANVDGRPLYAFGPNASYYESTGS
ASAPFRARTGSSVLFHAMQTYNHL"
complement(2812..2816)
/gene="ORF3"

RBS
BASE COUNT      486 a . 1192 c 1158 g 518 t
ORIGIN
```

Query Match 9.7%: Score 42: DB 73: Length 3354:

Best Local Similarity 44.9%: Pred. No. 11e+02:

Matches 159: Conservative 0: Mismatches 195: Indels 0: Gaps 0:

```
OY 9 cctcccggtcagcgccaccgcggctcctcctcccgagtttctgagctgttcgagc 68
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1810 CGTGGCGGAGCAGCTTTCGAGCTGACGCGCTCCCGGTCAATGATGCTCATGTACAC 1869
```

```
OY 69 cttcccgatctcgcgcggaactccgcccaccttcgaacaccggttgatgctgagaa 128
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1870 GTACCTCTTCGAGGGGCGCTGCGCGCTCGGTGACGGGTACATTCATTCTGCTGCC 1929
OY 129 cgaatgaagaagggcgctacgaagtacgcgcggaactcccgaggtcgaaccgacaa 188
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1930 CGGCATCTCTGATGAGCGCTGCTGATGATCAGATGTACACGGCGTCCGATCAACAC 1989
OY 189 ggaagtcacattatgctcgcgatgctgaagtcgaacatcaagggcgagcagagca 248
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1990 CGACATCGACAAAGGAGTCTTCGACCGCTTCGACAGCTGCCGATCTGGCGCGCCG 2049
OY 249 gaagacttaagcgttcgctcggaattcggtacggttcctcctcgttcgacaggtcgt 308
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 2050 GATGCTGGGCTATCTCTTCGCGGACGCTGCTCCGCTACTTCACTTCCTCGTGATGCT 2109
OY 309 gccgtaggtgctgacgaggaacacatcaagccaacctacgaaggagcatct 362
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 2110 GACGTCGGCGTCTGATCGGCTACCGCGCGGCGGCGGCGGACCGGATCTT 2163
```

Search completed: January 12, 2001, 20:08:02
Job time: 26656 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 5, 2001, 14:30:43 ; Search time 115.27 Seconds
(without alignments)
145.404 Million cell updates/sec

Title: US-09-461-774-8
Perfect score: 731
Sequence: 1 ATTLPRQRRPRSLFPEFSEL.....SVAVSGKPTREKHQIRSTN 143

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPREMBL_15:

1: sp-archaea:*
2: sp-bacteria:*
3: sp-fungi:*
4: sp-human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp-phage:*
10: sp-plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	165.5	22.6	159	2 053673	053673 mycobacteri
2	158	21.6	145	2 09XDC8	09XDC8 mycobacteri
3	154.5	21.1	143	3 014368	014368 schizosacch
4	149	20.4	146	2 P72977	P72977 synechocyst
5	148.5	20.3	169	2 09K4W0	09K4W0 stigmatella
6	147	20.1	140	1 028308	028308 archaeoglob
7	145.5	19.9	148	2 P94898	P94898 oenococcus
8	144.5	19.8	150	10 096458	096458 hordium vul
9	142.5	19.5	166	2 P70919	P70919 bradyrhizob
10	140.5	19.2	151	10 040056	040056 hordium vul
11	140	19.2	243	10 09SE11	09SE11 funaria hyg
12	138.5	18.9	142	2 09ZFP1	09ZFP1 thermotoga
13	137.5	18.8	163	2 069243	069243 bradyrhizob
14	137	18.7	159	10 040867	040867 pennisetum
15	136.5	18.7	145	2 082825	082825 synechococc
16	136.5	18.7	152	10 040866	040866 pennisetum
17	135.5	18.5	142	2 052192	052192 streptococc
18	135.5	18.5	154	2 09X9N3	09X9N3 streptococc
19	135	18.5	130	10 041561	041561 triticum ae

20	135	18.5	160	2 09PB80	09PB80 xylella fas
21	134.5	18.4	142	2 09X9C3	09X9C3 streptococc
22	133.5	18.3	150	2 052190	052190 streptococc
23	133.5	18.3	153	10 09ZSY1	09ZSY1 funaria hyg
24	133	18.2	147	2 09MYK7	09MYK7 thermotoga
25	133	18.2	151	10 041560	041560 triticum ae
26	133	18.2	142	2 09RVB5	09RVB5 delnocius
27	131.5	18.0	142	2 09RNV7	09RNV7 streptococc
28	131.5	18.0	142	2 09RNV9	09RNV9 streptococc
29	131.5	18.0	150	10 040865	040865 pennisetum
30	131.5	18.0	160	10 040057	040057 hordium vul
31	131	17.9	149	10 P93437	P93437 oryza sativ
32	131	17.9	210	5 040978	040978 papaver som
33	130	17.8	229	5 027354	027354 toxoplasma
34	129	17.6	152	10 043701	043701 zea mays (m
35	129	17.6	156	10 039930	039930 hellanthus
36	128	17.5	160	10 040632	040632 oryza sativ
37	127.5	17.4	154	2 09RNV3	09RNV3 streptococc
38	127.5	17.4	154	10 040698	040698 oryza sativ
39	127.5	17.4	154	10 09ZS24	09ZS24 castanea sa
40	127.5	17.4	154	10 09ZP84	09ZP84 quercus sub
41	127	17.4	157	10 022531	022531 brassica ra
42	127	17.4	160	10 P93440	P93440 oryza sativ
43	127	17.4	232	10 09SS08	09SS08 arabidopsi
44	127	17.4	238	10 09SE12	09SE12 funaria hyg
45	126.5	17.3	155	10 041028	041028 pisum sativ

ALIGNMENTS

RESULT 1

ID 053673 PRELIMINARY; PRT; 159 AA.

DI 01-JUN-1998 (TREMBLrel. 06, Created)
DI 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DI 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE PUTATIVE HSP20 HEAT SHOCK PROTEIN.
GN RV0251c OR MTW034.17C.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RM SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Gordon S.V., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Fellwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagers K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skellon S., Squares S., Squires R., Sulston J.E.,
RA Taylor K., Whitehead S., Barrett B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
DR EMBL: AL021929; CAA17343.1; -
DR TUBERCULIST: RV0251c; -
DR INTERPRO: IPR002068; -
DR PFAM: PF00011; HSP20; 1.
DR PROSITE: PS01031; HSP20; 1.
KW Heat shock.
SQ SEQUENCE 159 AA; 17786 MW; 6EB7DAFDE2C083CF CRC64;

Query Match 22.6%, Score 165.5; DB 2; Length 159;
Best Local Similarity 41.8%; Pred. No. 9.7e-08;
Matches 41; Conservative 13; Mismatches 33; Indels 11; Gaps 3;
Gy 52 VRAELPGVDPPKDVHIMRDQ-----LTKAKERTQK-----DLDRS--ERAVGSPVKT 100

DB 56 VLELPGIDVDKDVNVELDPCGVSRVLVIRGEHRDEHTODAGDKDRTILRIYSGSFRS 115
 QY 101 VSLPVGADDDIKATYDKGILTVSAVSESGKPTREKHQ 138
 DB 116 FRLPVHTSEALNASTYDAGVLTVRAGAVKAPAEQAQ 153

RESULT 2

Q9XDC8 PRELIMINARY: PRT: 145 AA.

AC 09XDC8.
 DT 01-NOV-1999 (TREMBlrel. 12, Created)
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
 DE PUTATIVE HSP20.
 GN HSP.
 OS Mycobacterium smegmatis.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacteriaceae; Mycobacterium.

NCBI_TaxID=1772;
 [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=MC(2)15;
 RA Caceres N.E., Feng Z., Li L.-L., Kapur V., Cirillo J.D., Barletta R.G.;
 RT "Identification of a HSP20 heat shock protein homolog from Mycobacterium smegmatis."
 RL Submitted (JUN-1998) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AF072875; AAD39038.1;
 DR HSP: Q57733; ISHS.
 DR INTERPRO: IPR002068;
 DR PFM: PF00011; HSP20; 1.
 DR PROSITE: PS01031; HSP20; 1.
 SQ SEQUENCE 145 AA; 16202 MW; 978DC47F74BE30E9 CRC64;

Query Match 21.6%; Score 158; DB 2; Length 145;
 Best Local Similarity 42.6%; Pred. No. 4.2e-07;
 Matches 40; Conservative 12; Mismatches 38; Indels 4; Gaps 2;

QY 52 VNAELPGVDKDVHIMVWDGOLITKAERTECK--DLGR--SEFAYGSFVRTSLPVG 107
 DB 51 VVELPGVDVGVNVEVDRGLVHGERDRSESDGRTYSEVAYGSFRREKLPVAV 110
 QY 108 DEDDIKATYDKGILTVSAVSESGKPTREKHQIR 141
 111 TGDVASTYDAGVLTVRAGAVKAPAEQAQ 144

RESULT 3

Q14368 PRELIMINARY: PRT: 143 AA.

AC 014368.
 DT 01-JAN-1998 (TREMBlrel. 05, Created)
 DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
 DE HEAT SHOCK PROTEIN 16.
 GN HSP16.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetales; Schizosaccharomycetaceae; Schizosaccharomycetes.
 NCBI_TaxID=4896;
 [1]

RP SEQUENCE FROM N.A.

RC STRAIN=972 H-;
 RA Lenaers G., Perret E., Dumont X., Picard A., Caput D.;
 RL Submitted (OCT-1997) to the EMBL/Genbank/DBJ databases.
 [2]

RP SEQUENCE FROM N.A.
 RC STRAIN=972H-;
 RA Lyne M., Wood V., Rajandream M.A., Barrell B.G., Brown D., Churcher C.M.;

RL Submitted (MAY-1998) to the EMBL/Genbank/DBJ databases.
 RN [3]

RP SEQUENCE FROM N.A.
 RA Danjoh I., Fujiyama A.;
 RT "Ras-mediated signaling pathway regulates the expression of a low molecular weight heat shock protein in fission yeast."
 RL Submitted (MAR-1998) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AJ003817; CA06031.1;
 DR EMBL: AL023534; CA019006.1;
 DR EMBL: AB012619; BAA31521.1;
 DR HSP: Q57733; ISHS.
 DR INTERPRO: IPR002068;
 DR PFM: PF00011; HSP20; 1.
 DR PROSITE: PS01031; HSP20; 1.
 KW Heat shock.

SQ SEQUENCE 143 AA; 15968 MW; 321D474116FCE380 CRC64;

Query Match 21.1%; Score 154.5; DB 3; Length 143;
 Best Local Similarity 32.2%; Pred. No. 8.5e-07;
 Matches 46; Conservative 25; Mismatches 45; Indels 27; Gaps 6;

QY 15 PEESELPAAPPSFA-----GLRPTDPTRLRLDEMKR--YEVRAELPGVDPK 63
 DB 11 PTVNDLFSDPVSXSPRLNNQIPGLSPSID-----VHEGKTVSYDVELPGVKE- 60
 QY 64 DVHIMVWDGOLITKAERTECKDLGR--SEFAYGSFVRTSLPVGADDDIKATYD 117
 DB 61 DVQVHSDSKLTISGEVVERKNESTEGNQRWSERRGFSFRTITPAKIDARLEANS 120
 QY 118 KGLTVSAVSESGKPTREKHQIR 140
 DB 121 NGLITVTLPEKVESQTKQIAIK 143

RESULT 4

P72977 PRELIMINARY: PRT: 146 AA.

AC P72977.
 DT 01-FEB-1997 (TREMBlrel. 02, Created)
 DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)
 DT 01-JUN-2000 (TREMBlrel. 14, Last annotation update)
 DE SPORE PROTEIN SP21.
 GN HSPA.
 OS Synechocystis sp. (strain PCC 6803).
 OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.

NCBI_TaxID=1148;
 [1]

RP SEQUENCE FROM N.A.
 RX MEDLINE=97061201; PubMed=8905231;
 RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y., Miyajima N., Hirosewa M., Sugiyura M., Sasamoto S., Kimura T., Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naito K., Okumura S., Shimo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Tabata S.;
 RT "Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions."
 RL DNA Res. 3:109-136(1996).
 DR EMBL: D90902; BAA16996.1;
 DR INTERPRO: IPR002068;
 DR PFM: PF00011; HSP20; 1.
 DR PROSITE: PS01031; HSP20; 1.
 SQ SEQUENCE 146 AA; 16596 MW; 4453B15DF8F9EABE CRC64;

Query Match 20.4%; Score 149; DB 2; Length 146;
 Best Local Similarity 36.7%; Pred. No. 2.7e-06;
 Matches 36; Conservative 19; Mismatches 37; Indels 6; Gaps 4;

QY 47 EGRYEVRAELPVDPKDVHIMVWDGOLITKAER--TECKDLGR--RSEFAYGSFVRTS 102
 DB 49 EEAAYLKLLEPMDPD-NLDIQARDVAVSGDRQDHTSTKDGVRRTFRGSGFRVLP 107

DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, last sequence update)
 DT 01-JUN-2000 (TREMBLrel. 14, last annotation update)
 DE CHLOROPLAST-LOCALIZED SMALL HEAT SHOCK PROTEIN 22.
 GN CPSSSP22.
 OS Funaria hygrometrica.
 OC Eukaryota; Viridiplantae; Embryophyta; Bryophyta; Bryopsida;
 OC Funariidae; Funariales; Funariaceae; Funaria.
 OX NCBI_TaxID=29583;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Waters E.R., Vierling E.;
 RT "Chloroplast small heat shock proteins: Evidence for atypical
 evolution of an organelle-localized protein."
 RL Proc. Natl. Acad. Sci. U.S.A. 96:14394-14399(1999).
 DR EMBL; AF197942; AAF19022.1; -.
 DR INTERPRO; IPR002068; -.
 DR PFAM; PF00011; HSP20; 1.
 KM Heat shock.
 SEQUENCE 243 AA; 27468 MW; EA45002D4C1B690B CRC64;

Query Match 19.2%; Score 140; DB 10; Length 243;
 Best Local Similarity 33.6%; Pred. No. 3.4e-05;
 Matches 41; Conservative 23; Mismatches 50; Indels 8; Gaps 4;

OY 21 FAAFSFGALRTPTDRLRLDEDEKGR-YEVRAELPGVPDGDVHIMVADGQLTIK- 77
 DB 124 FSAAPS---RPTGIALDPTPMDVKEDNESFRLRDMGLKD-EVKYVBDGDIYVIG 178
 OY 78 AERTGOKLDGSEFASVFRTVSLPGVADDDIKATYDKGILTVSAVSEGKPTKEHI 137
 DB 179 AHRAEOKKNNSSSYGSYNTRMTLPENVKIDEVAKELNGVLYVVKSEEPKKNVI 238
 OY 138 QI 139
 DB 239 DI 240

RESULT 12
 O9ZFD1 PRELIMINARY; PRT; 142 AA.
 AC O9ZFD1;
 DT 01-MAY-1999 (TREMBLrel. 10, Created)
 DT 01-MAY-1999 (TREMBLrel. 10, last sequence update)
 DT 01-MAY-2000 (TREMBLrel. 13, last annotation update)
 DE SMALL HEAT SHOCK PROTEIN.
 SHSP.
 OS Thermotoga maritima.
 OC Bacteria; Thermotogales; Thermotoga.
 OX NCBI_TaxID=2336;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Michellini E.T., Flynn G.C.;
 RT "The unique chaparrone operon of Thermotoga maritima: cloning and
 initial characterization of a functional Hsp70 and shsp."
 RL Submitted (NOV-1998) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AF106330; AAC79726.1; -.
 DR INTERPRO; IPR002068; -.
 DR PFAM; PF00011; HSP20; 1.
 DR PROSITE; PS01031; HSP20; 1.
 KM Heat shock.
 SEQUENCE 142 AA; 17006 MW; D7B6B0C75D08F565 CRC64;

Query Match 18.9%; Score 138.5; DB 2; Length 142;
 Best Local Similarity 29.7%; Pred. No. 2.4e-05;
 Matches 41; Conservative 25; Mismatches 53; Indels 19; Gaps 5;
 OY 6 VORHPSLPPESEFLFAAFPSFAGLPTDRLRLDEDEKGRYVRAELPGVDPDKDV 65
 DB 16 LQREIDRLFDDEFR-----TEVRPAPDMDFETDDEV-----VIEVEIPCID-RKDV 61

OY 66 HIMVBDGQLTIKAERTGOKLDGRS---EFAYGSFVRTSLPGVADDDIKATYDKGIL 121
 DB 62 QITVEENILIKISGEKKELEDEOKKNYYERSAGFERRAIRLDPYVDVKIKAIKRNGL 121
 OY 122 TVSAVSEGKPTKEHIQI 139
 DB 122 TIRVPKKEER-KKKYIEV 138

RESULT 13
 O69243 PRELIMINARY; PRT; 163 AA.
 AC O69243;
 DT 01-AUG-1998 (TREMBLrel. 07, Created)
 DT 01-AUG-1998 (TREMBLrel. 07, last sequence update)
 DT 01-MAY-2000 (TREMBLrel. 13, last annotation update)
 DE HSPF.
 GN HSPF.
 OS Bradyrhizobium japonicum.
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Bradyrhizobium group; Bradyrhizobium.
 OX NCBI_TaxID=375;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-96283699; PubMed-9622356;
 RA Narberhaus F., Kaeser R., Nocker A., Hennecke H.;
 RT "A novel DNA element that controls bacterial heat shock gene
 expression."
 RL Mol. Microbiol. 28:315-323(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-96404779; PubMed-8808920;
 RA Narberhaus F., Weighlofer W., Fischer H.M., Hennecke H.;
 RT "The Bradyrhizobium japonicum rpoH gene encoding a sigma 32-like
 protein is part of a unique heat shock gene cluster together with
 rpoES1 and three small heat shock genes."
 RL J. Bacteriol. 178:5337-5346(1996).
 DR EMBL; AJ003064; CA05837.1; -.
 DR INTERPRO; IPR002068; -.
 DR PFAM; PF00011; HSP20; 1.
 DR PROSITE; PS01031; HSP20; 1.
 SQ SEQUENCE 163 AA; 18645 MW; 6EB5CAACE06028E CRC64;

Query Match 18.8%; Score 137.5; DB 2; Length 163;
 Best Local Similarity 43.3%; Pred. No. 3.5e-05;
 Matches 42; Conservative 16; Mismatches 28; Indels 11; Gaps 7;

OY 44 EMKEGEYVR-AELPGVPDGDVHIMVADGQLTIKAE-RTGOKLDGR-SEFAYGSFVR 99
 DB 62 ERENEDKEVKVIAELPGIE-QKDVEVELADGVLTISGEKSETKDKERFSERYTGRFR 120
 OY 100 TVSLPV-GAEDDDIKATYDKGILTVSAVSEGKPTKEK 135
 DB 121 RI-PYDDVDGQKVASFKDGLITVTLPS---PTFO 152

RESULT 14
 O40867 PRELIMINARY; PRT; 159 AA.
 AC O40867;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)
 DT 01-MAY-2000 (TREMBLrel. 13, last annotation update)
 DE HEAT SHOCK PROTEIN 17.9.
 GN HSP17.9.
 OS Penicillium americanum (pearl millet).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; Liliopsida; Poales; Poaceae; Pennisetum.
 OX NCBI_TaxID=4543;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-LRE179;

RA Cavan G.;
RL Submitted (Dec-1995) to the EMBL/GenBank/DBJ databases.
RP [2]
RQ SEQUENCE FROM N.A.

RX MEDLINE=97156146; PubMed=9002606;
RY Caven G.P., Skot K., Stevens M.J., Howarth C.J.;
RT "Sequence homology requirements for transcriptional silencing of 35S
RT transgenes and post-transcriptional silencing of nitrite reductase
RT (trans)genes by the tobacco 271 locus";
RT Plant Mol. Biol. 30:1075-1075(1996).
DR EMBL: X94193; CAA63903.1; -;
DR MENDEL: 97519; Pengl:1195;9759.
DR INTERPRO: IPR002068; -;
DR PRAM: PF000011; HSP20_1
KW PROSITE: PS01031; HSP20_1.
KW Heat Shock.
SQ SEQUENCE 159 AA; 17946 MW; C308EF7E15126C03 CRC64;

Every Match	18.7%	Score 137;	DB 10;	Length 159;
Best Local Similarity	32.8%	Pred. No. 3.8e-05;		
Matches 44;	Conservative 21;	Mismatches 53;	Indels 16;	Gaps 5;

```

QY 12 SLPEESELFAFPSPGLRPTFTRLMLRLDEDEKKEGYEVRALPLVPDDEPKDVHIVARD 71
    ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 34 SLPSFPSTSETSAFAFG-----ARIDWKETTPEAHVAKADVPALPKE-EYKVEVED 83

QY 72 GQ-FLTKAERT----EKKDLGRSEFAVSPVRYVSLPYGDEEDIKATYDKGLITYSVA 126
    ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 84 GNVLTQISGRKKDEDEKTDTHWRVERSSGKFRFRFLPENAKTDQITASHMENGVLIVTP 143

QY 127 VSE-GKPTKHIQT 139
    ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 144 KEVKKKEPVKSIQT 157
    ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||

```

RESULT	ID	PRELIMINARY;	PRT;	145 AA.
082825	ID 082825			
AC 082825;				
DT 01-NOV-1998 (TrEMBLrel. 08, Created)				
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)				
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)				
DE HSPA.				
GN HSPA.				
OS <i>Synechococcus vulcanus</i> .				
OS <i>Bacteria; Cyanobacteria; Chroococcales; Synechococcus</i> .				
NCBI_TaxID=32053;				
[1]				
SEQUENCE FROM N.A.				
RP MEDLINE=98348466; PubMed=9683501;				
RA Roy S.K., Nakamoto H.;				
RT "Cloning, characterization, and transcriptional analysis of a gene encoding an alpha-crystallin-related, small heat shock protein from the thermophilic cyanobacterium <i>Synechococcus vulcanus</i> ."				
RT J. Bacteriol. 180:3997-4001(1998).				
EMBL; AB002666; BAA32501.1; -				
DR INTERPRO; IPR002068; -				
DR PRAM; PF00011; HSP20.1.				
DR PROSITE; PS01031; HSP20.1.				
Q0 SEQUENCE 145 AA; 16650 MW; D21B30C1CBACE956 CRC64;				

Query Match	18.7%;	Score 136.5;	DB 2;	Length 145;
Best Local Similarity	30.8%;	Pred. No. 3.7e-05;		
Matches 41; Conservative	21;	Mismatches 48;	Indels 23;	Gaps 5;

Qy 6 VQRPRRLPEEFSLL-----AAPSPAGLRPPFDOTLMLLEDBMKGRTEVAALRPGVD 60
::| | | :| | :|||:
Db 15 IQRMKNRLFDLLPLTRRSDSLTPAALEEIPALLKY-----ELPQMD 61

Qy -, 61 PDKDVHLMVADGQLTIAERTEQ--KLDG--ASEAFSGSVFRTVSLPGCADEDIKATY 116

Db 62 P-KDIDOVTAEAVSISERSKSEKTETEGMKRTFEYRGAFVPIPLVRLONTSYAEV 120
 QY 117 DKGILTVSAVSE 129
 Db 121 KDGIHLHLTPKAE 133

Search completed: January 5, 2001, 14:30:45
Job time: 337 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 5, 2001, 14:31:21 ; Search time 35.06 Seconds

(Without alignments)
131.719 Million cell updates/sec

Title: US-09-461-774-8

Perfect score: 731
Sequence: 1 ATLLPYQRHRSLEFPPESEL.....SVAVSCKPTKEKHQIRSTN 143

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 88757 seqs, 32294092 residues

Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SWISSProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	718	98.2	143	14KD_MYCTU	P30223 mycobacteri
2	148.5	20.3	188	SP21_STIAU	O06823 stigmatella
3	137.5	18.8	157	HS2C_CHLRE	P12811 chlamydomon
4	130.5	17.9	142	ASPI1_STRTR	P80485 streptococ
5	130.5	17.9	159	HS12_DAUCA	P27397 daucus caro
6	130	17.8	151	HS11_WHEAT	P17810 triticum ae
7	129	17.6	153	HS11_HELAN	P30693 helianthus
8	127.5	17.4	161	HS16_SOYBN	P05478 glycine max
9	126.5	17.3	161	HS11_CHERU	O05832 chenopodium
10	126	17.2	213	HS26_YEAST	P15992 saccharomyc
11	124.5	17.0	154	HS12_ORYSA	P16773 oryza sativ
12	124	17.0	156	HS12_ARATH	P13853 arabidopsis
13	123.5	16.9	156	HS11_ARATH	P19036 arabidopsis
14	123.5	16.9	161	HS13_ARATH	P19037 arabidopsis
15	121.5	16.6	150	HS11_ORYSA	P27777 oryza sativ
16	121	16.6	143	HS18_STRAL	O53595 streptomyc
17	119.5	16.3	235	HS2C_LYCSC	O95661 lycopersico
18	119.5	16.3	241	HS2C_PETHY	P30222 petunia hyb
19	119	16.3	143	HS11_MEDSA	P27879 medicago sa
20	118.5	16.2	197	HS41_PEA	P19244 pisum sativ
21	118	16.1	153	HS11_SOYBN	P02519 glycine max
22	117.5	16.1	158	HS11_PEA	P19243 pisum sativ
23	117	16.0	142	ASR2_STRTR	O36851 streptococ
24	116	15.9	154	HS11_LYCSC	P30221 lycopersico
25	116	15.9	158	HS12_MEDSA	P27880 medicago sa
26	113.5	15.5	181	HS30_EMENI	P40920 emericella
27	112.5	15.4	227	HS14_SOYBN	P04794 arabidopsis
28	112	15.3	227	HS2C_ARATH	P31170 arabidopsis
29	110.5	15.1	153	HS13_SOYBN	P04793 glycine max
30	109.5	15.0	157	HS11_DAUCA	P27396 daucus caro
31	109	14.9	137	HS2C_PEA	P09886 pisum sativ
32	106.5	14.6	138	18K2_MYCAV	P46731 mycobacteri
33	106	14.5	154	HS15_SOYBN	P04795 glycine max

34	104.5	14.3	228	1	HS30_NEUCR	P19752 neurospora
35	104	14.2	161	1	HS21_MAZE	P24631 zea mays (m
36	103	14.1	160	1	HS21_HELAN	P46516 helianthus
37	102.5	14.0	149	1	18K1_MYCT	P46730 mycobacteri
38	102	14.0	154	1	HS23_MAZE	O08275 zea mays (m
39	102	14.0	167	1	HS22_PMANI	O01545 pharbitis n
40	102	14.0	192	1	HS41_SOYBN	P30236 glycine max
41	101.5	13.9	148	1	18KD_MYCLE	P12809 mycobacteri
42	101.5	13.9	245	1	ODFP_RAT	P21769 ratius norv
43	101.5	13.9	247	1	ODFP_MOUSE	O61999 mus musculu
44	101.5	13.9	250	1	ODFP_HUMAN	O14990 homo sapien
45	101.5	13.9	262	1	ODFP_BOVIN	O29438 bos taurus

ALIGNMENTS

```

RESULT 1
14KD_MYCTU STANDARD: PRT: 143 AA.
AC P30223;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE 14 KDA ANTIGEN (16 KDA ANTIGEN) (HSP 16.3).
GN HSPX OR RV2031C OR MTV018.18C.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ERDMANN, AND H37RV;
RX MEDLINE=92138631; PubMed=1370952;
RA Verdon A., Hartskeerl R.A., Schultema A., Kolk A.H.J., Young D.B.,
RA Lathigra R.;
RT "The 14,000-molecular-weight antigen of Mycobacterium tuberculosis is
RT related to the alpha-crystallin family of low-molecular-weight heat
RT shock proteins."
RL J. Bacteriol. 174:1352-1359(1992).
RN [2]
RP SEQUENCE, AND CHARACTERIZATION.
RC STRAIN-ERDMANN;
RX MEDLINE=92225631; PubMed=1563797;
RA Lee B.-Y., Hetta S.A., Brennan P.J.;
RT "Characterization of the major membrane protein of virulent
RT Mycobacterium tuberculosis."
RL Infect. Immun. 60:2066-2074(1992).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-H37RV;
RX MEDLINE=96295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltham D., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sultson J.E.,
RA Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
CC -1- SUBCELLULAR LOCATION: PROBABLY THE EXTERNAL SIDE OF THE CELL WALL.
CC -1- MASS SPECTROMETRY: MW=16100; METHOD=ELECTROSPRAY.
CC -1- SIMILARITY: BELONGS TO THE SMALL HEAT SHOCK PROTEIN (HSP20)
CC FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation-
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/

```



```

Query Match      18.8%; Score 137.5; DB 1; Length 157;
Best Local Similarity 38.8%; Pred. No. 6.9e-06;
Matches 38; Conservative 14; Mismatches 37; Indels 9; Gaps 4;

QY 50 YEVRALPGVDPDKVHIMVWDGQLTIKAERT---EOKDLG--RSEFAYGSFVRYSL 103
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 58 FELHADAGMGPD-DVKVELQEGVLMTGGERKLSHTTKKAGKQWRSRTAYSFGRSL 116
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 104 PVGADDEDIKATYDKGILTVSAVSE--GKPTKHIQI 139
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 117 PENAMPDGIATAMDKGVLTIVYKREPPAKPEKRIAV 154
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 4
ASPL STRR
ID ASPL_STRR STANDARD; PRT; 142 AA.
AC P80485; P80914;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
ACID SHOCK PROTEIN (T786P28D).
Strepococcus thermophilus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
Strepococcus.
RN [1]
RP STRAIN=PB18;
RC Gonzalez-Marquez H., Bracquart P., Guilmont C., Linden G.,
RA van Dorsseleir A., Sorokine O.;
RT "Identification and characterization of an acid shock protein in
Strepococcus thermophilus.";
RL Submitted (FEB-1997) to the SWISS-PROT data bank.
RN [2]
RP SEQUENCE OF 1-10.
RC STRAIN=PB18;
RX MEDLINE=97311999; Pubmed=9168610;
RA Gonzalez-Marquez H., Perrin C., Bracquart P., Guilmont C., Linden G.;
RT "A 16 kDa protein family overexpressed by Streptococcus thermophilus
in acid environments.";
RL Microbiology 143:1587-1594(1997).
CC -1- INDUCTION: OVEREXPRESSED IN ACID ENVIRONMENTS.
CC -1- SIMILARITY: BELONGS TO THE SMALL HEAT SHOCK PROTEIN (HSP20)
FAMILY.
DR INTERPRO: IPR002068;
DR PFAM: PF00011; HSP20; 1.
DR PROSITE: PS01031; HSP20; 1.
SQ SEQUENCE 142 AA; 16442 MM; 0B57D9AEF8030BA CRC64;

Query Match      17.9%; Score 130.5; DB 1; Length 142;
Best Local Similarity 29.3%; Pred. No. 2.7e-05;
Matches 41; Conservative 22; Mismatches 52; Indels 25; Gaps 6;

QY 10 PRSLPEEPS-ELFAFPSEAGLRPFQDRLMRLEDEMKEGREVARELPGVDPDKVHIM 68
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 17 PDPEEPEESRLNDF-----KPNF---IKTDIHENDNEIYVAELPGI-PKRNIOYT 65
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 69 VRDGLRTIKARTEQKDLGRSEFAYGSFVRTVS-----LFGADEDIDIKATYDKG 119
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 66 YENGVLITSG---QQQIDAVNEDKKGLINSERSLTSVQRYLLLENKVEDEIKASYS DG 121
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 120 ILTVSVANSEKPTKHIQI 139
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 122 VLKVTLPKDSNKEIKKISISI 141
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 5
HS12 DAUCA STANDARD; PRT; 159 AA.
ID HS12_DAUCA
AC P27397;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)

```

```

DE 18.0 KDA CLASS I HEAT SHOCK PROTEIN (CLONE DCHSP17.9).
OS Eucarya carota (Carrot).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Asteridae;
OC euasterids II; Apiales; Apiaceae; Daucus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. DANVERS HALF-LONG;
RX MEDLINE=91329706; Pubmed=1840684;
RA "Wang L., Wang L., Huang C.H., Apuya N., Zimmerman J.L.;
RT "Cloning and characterization of genes encoding low molecular weight
heat shock proteins from carrot.";
RL Plant Mol. Biol. 16:729-731(1991).
CC -1- SUBUNIT: FORMS OLIGOMERIC STRUCTURES.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SIMILARITY: BELONGS TO THE SMALL HEAT SHOCK PROTEIN (HSP20)
FAMILY.
CC -1- SIMILARITY: PLANTS HAVE FOUR CLASSES OF HSP20: CLASS I AND II
WHICH ARE CYTOPLASMIC; CLASS III WHICH IS IN THE CHOROPLAST
AND CLASS IV WHICH IS IN THE ENOMEMBRANE. THIS PROTEIN BELONGS
TO CLASS I.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@lsb-sib.ch).
CC -----
DR EMBL: X53852; CAA37848.1;
DR PIR: S15000; CYP279.
DR INTERPRO: IPR002068;
DR PFAM: PF00011; HSP20; 1.
DR PROSITE: PS01031; HSP20; 1.
KW Heat shock; Multigene family.
SQ SEQUENCE 159 AA; 18032 MM; 46B004F4921D64B8 CRC64;

Query Match      17.9%; Score 130.5; DB 1; Length 159;
Best Local Similarity 37.6%; Pred. No. 3e-05;
Matches 35; Conservative 19; Mismatches 32; Indels 7; Gaps 4;

QY 53 RAELPGVDPDKVHIMVWDGQ-LTIKAERTDQKIDG---RSEFAYGSFVRYSLPVGA 107
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 66 KADLPGLKKE-EVKVEVEGEGVQLQISGERNKEKKNKMRVERFSGKFLRRFLPEN A 124
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 108 DEDDIKATYDKGILTVSY-AVSEKPTKHIQI 139
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 125 NVDEVKAGMENGVLTIVYKVEKMKPEVKSITHI 157
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 6
HS11 WHEAT STANDARD; PRT; 151 AA.
ID HS11_WHEAT
AC P12810;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE 16.9 KDA CLASS I HEAT SHOCK PROTEIN (LOW MOLECULAR WEIGHT HEAT SHOCK
DE PROTEIN) (HEAT SHOCK PROTEIN 17) (HSP 16.9).
OS Triticum aestivum (Wheat).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; Liliopsida; Poales; Poaceae; Triticum.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. YAMHILL;
RX MEDLINE=89160335; Pubmed=2922294;
RA McElwain E.F., Spiker S.;
RT "A wheat cDNA clone which is homologous to the 17 kd heat-shock
protein gene family of soybean.";
RL Nucleic Acids Res. 17:1764-1764(1989).
CC -1- SUBUNIT: FORMS OLIGOMERIC STRUCTURES.

```

```
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SIMILARITY: BELONGS TO THE SMALL HEAT SHOCK PROTEIN (HSP20)
CC FAMILY.
CC -1- SIMILARITY: PLANTS HAVE FOUR CLASSES OF HSP20: CLASS I AND II
CC WHICH ARE CYTOPLASMIC; CLASS III WHICH IS IN THE CHLOROPLAST
CC AND CLASS IV WHICH IS IN THE ENDOMEMBRANE. THIS PROTEIN BELONGS
CC TO CLASS I.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X13431; CAA31785.1; -
CC PIR: S03178; HHMT17.
CC INTERPRO: IPR002068; -
CC PFAM: PF00011; HSP20; 1.
CC PROSITE: PS01031; HSP20; 1.
CC Heat shock; Multigene family.
CC SEQUENCE 151 AA; 16878 MW; E0D6DBA3CB30B310 CRC64;
SQ
Query Match 17.8%; Score 130; DB 1; Length 151;
Best Local Similarity 30.8%; Pred. No. 3.2e-05;
Matches 44; Conservative 29; Mismatches 54; Indels 16; Gaps 7;
QY 12 SLPFSESEFA-AFPSPAGLRPTF-----DTRLM---RLEDEMKEGREYKRALPGVDDP 62
DB 8 NFDPPADIMADPFDFFRSIVPAISGSSSETAFANARVDMKETPRAHFKYDLPGVKE 67
QY 63 KDVIHWBGOQ-LTIAERTECK---DLDRGEFAYGVSVTVSLPYGADDEDIKATYD 117
DB 68 -EKVVEEGGNLVVSGENREKEDKNDKHWYRVSNGKFRRLPEDAKYEVRAGLE 126
QY 118 KGLTVSAVSE-GKPRTEKHQI 139
DB 127 NGVLTVTPKAEYKRPVKATEI 149
RESULT 7
HS11_HELAN STANDARD; PRT; 153 AA.
AC P30693;
DT 01-APR-1993 (Rel. 25, Created)
01-APR-1993 (Rel. 25, Last sequence update)
15-JUL-1998 (Rel. 36, Last annotation update)
17.6 KDA CLASS I HEAT SHOCK PROTEIN.
GN HSP17.6.
OS Helianthus annuus (Common sunflower).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Asteridae;
OC Euasterids II; Asterales; Asteraceae; Asteroideae; Heliantheae;
OC Helianthus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. SUNMEED; TISSUE=SEED;
RX MEDLINE=92353387; PubMed=1386536;
RA Almoguera C., Jordano J.;
RT "Developmental and environmental concurrent expression of sunflower
RT dry-dweld-stored low-molecular-weight heat-shock protein and Lea
RT mRNAs";
RL Plant Mol. Biol. 19:781-792(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=98001736; PubMed=9341201;
RA Carranco R., Almoguera C., Jordano J.;
RT "A plant small heat shock protein gene expressed during zygotic
RT embryogenesis but noninducible by heat stress.";
RL J. Biol. Chem. 272:27470-27475(1997).
CC -1- SUBUNIT: FORMS OLIGOMERIC STRUCTURES.
```

```
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SIMILARITY: BELONGS TO THE SMALL HEAT SHOCK PROTEIN (HSP20)
CC FAMILY.
CC -1- SIMILARITY: PLANTS HAVE FOUR CLASSES OF HSP20: CLASS I AND II
CC WHICH ARE CYTOPLASMIC; CLASS III WHICH IS IN THE CHLOROPLAST
CC AND CLASS IV WHICH IS IN THE ENDOMEMBRANE. THIS PROTEIN BELONGS
CC TO CLASS I.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X59701; CAA42222.1; -
CC PIR: S25153; CAB08441.1; -
CC PIR: S18166; S18166.
CC PIR: S23529; S23529.
CC INTERPRO: IPR002068; -
CC PFAM: PF00011; HSP20; 1.
CC PROSITE: PS01031; HSP20; 1.
CC Heat shock; Multigene family.
CC SEQUENCE 153 AA; 17562 MW; DAC8FC46FD253C3C CRC64;
SQ
Query Match 17.6%; Score 129; DB 1; Length 153;
Best Local Similarity 31.7%; Pred. No. 4e-05;
Matches 44; Conservative 26; Mismatches 57; Indels 12; Gaps 7;
QY 12 SLPFERS-ELFAFPSPAGLRPTFDTRLM---RLEDEMKEGREYKRALPGVDDPKVHL 67
DB 14 NFDPPSLDTWPFQGITSTPARETAIYNRIQKTEPNAHYLKAADLPQKKKE-EVAV 72
QY 68 MVRDGO-LTIAER-TEQKDLG---RSEFAYGSEVTVSLPYGADDEDIKATYDKILT 122
DB 73 EVEDGRLVQISGERCREQEKDDTWHYRVSNGKFRRLPENAKMDEVKAMENGVL 132
QY 123 VSY-AVSEGKPRTEKHQI 139
DB 133 VVVPKEEERKKPVKAI 151
RESULT 8
HS16_SOYBN STANDARD; PRT; 161 AA.
AC P05478;
DT 01-NOV-1988 (Rel. 09, Created)
01-NOV-1988 (Rel. 09, Last sequence update)
01-FEB-1994 (Rel. 28, Last annotation update)
18.5 KDA CLASS I HEAT SHOCK PROTEIN (HSP 18.5).
GN HSP18.5-C.
OS Glycine max (Soybean).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
OC Fabales; Fabaceae; Papilionoideae; Glycine.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. CORSOY;
RX MEDLINE=88172502; PubMed=3351943;
RA Raschke E., Baumann G., Schoeffl F.;
RT "Nucleotide sequence analysis of soybean small heat shock protein
RT genes belonging to two different multigene families.";
RL J. Mol. Biol. 199:549-557(1988).
RN [2]
RP -1- SUBUNIT: FORMS OLIGOMERIC STRUCTURES.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SIMILARITY: BELONGS TO THE SMALL HEAT SHOCK PROTEIN (HSP20)
CC FAMILY.
CC -1- SIMILARITY: PLANTS HAVE FOUR CLASSES OF HSP20: CLASS I AND II
CC WHICH ARE CYTOPLASMIC; CLASS III WHICH IS IN THE CHLOROPLAST
CC AND CLASS IV WHICH IS IN THE ENDOMEMBRANE. THIS PROTEIN BELONGS
CC TO CLASS I.
```



```

DR   PIR; S20803; S20803.
DR   PIR; S33566; S33566.
DR   INTERPRO; IPR002068; -.
DR   Pfam; PF00011; HSP20; 1.
DR   PROSITE; PS01031; HSP20; 1.
KW   Heat shock.
SQ   SEQUENCE   161 AA;  18271 MW;  D9787AFAD66AD66 CRC64;

Query Match          17.3%; Score 126.5; DB 1; Length 161;
Best Local Similarity 30.1%, Pred. No. 7.1e-05;
Matches 44; Conservative 23; Mismatches 60; Indels 19; Gaps 6;

QY   12 SLEPERS-----ELFAFPSPFAGLRLPTFDTRL-----MRLEDMKGRYEVRALPGV 59'
      ::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB   16 NIFPFSSDELDWDPFFGLPSTLSTVPRSRETAETAAAFANARIDWKETPEAHVFKADLPGV 75'
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY   60 DPKADVHIMVNDGQ-LTIKAERTEK-----DLDGSEFAYGSFVATVSLPGVADDDLIKA 114
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB   76 KKE-EVKEVEVDGNLRTISGQARAREKKNPTWHRVERSSGQFMKFRLLPENAKVDQVKA 134
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY   115 TYDKGILTVSAVASEG-KPTEKHIOI 139
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB   135 GMENCVLTIVTPVKNKNEAPKPYKAINV 160
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

RESULT 10
HS26_YEAST
ID   HS26_YEAST          STANDARD;          PRT;    213 AA.
AC   P15992;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-NOV-1990 (Rel. 16, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   HEAT SHOCK PROTEIN 26 (26 KDA HEAT SHOCK PROTEIN).
GN   HSP26 OR YBR072W OR YBR0714.
GN   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC   Saccharomycetaceae; Saccharomyces.
NC   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90097950; Pubmed=2689876;
RA   Susek R.E., Lindquist S.L.;
RT   "hsp26 of Saccharomyces cerevisiae is related to the superfamily of
RT   small heat shock proteins but is without a demonstrable function.";
RL   Mol. Cell. Biol. 9:5265-5271(1989).
NC   [2]
RN   SEQUENCE FROM N.A.
RX   MEDLINE=89378758; Pubmed=2673926;
RA   Bossier P., Fitch I.T., Boucherie H., Tuile M.F.;
RT   "Structure and expression of a yeast gene encoding the small
RT   heat-shock protein Hsp26.";
RL   Gene 78:323-330(1989).
NC   [3]
RN   SEQUENCE FROM N.A.
RP   STRAIN=S288C;
RX   MEDLINE=95076715; Pubmed=7985423;
RA   van der Aart O.J.M., Barthe C., Doignon F., Aigle M., Crouzet M.,
RA   Steensma H.Y.;
RT   "Sequence analysis of a 31 kb DNA fragment from the right arm of
RT   Saccharomyces cerevisiae chromosome II.";
RL   Yeast 10:959-964(1994).
CC   -1- FUNCTION: NOT KNOWN. ONE OF THE MAJOR POLYPEPTIDES PRODUCED ON
CC   HEAT SHOCK.
CC   -1- SUBUNIT: PRESENT IN LARGE COMPLEXES.
CC   -1- DEVELOPMENTAL STAGE: EXPRESSED DURING THE ENTRY INTO STATIONARY
CC   PHASE RESULTING FROM GLUCOSE LIMITATION.
CC   -1- INDUCTION: BY HEAT SHOCK, AND UNDER OTHER CONDITIONS OF STRESS,
CC   SUCH AS INCREASED SALT CONCENTRATION AND STARVATION.
CC   -1- SIMILARITY: BELONGS TO THE SMALL HEAT SHOCK PROTEIN (HSP20)
CC   FAMILY.
CC   -----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL Outstation -

```

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL, M23871; AAA66914.1; -
DR EMBL, M26942; AAA79010.1; -
DR EMBL, X76294; CAA53929.1; -
DR EMBL, Z35941; CAA85016.1; -
DR PIR, S45465; S45465.
DR PIR, S39222; S39222.
DR SGD, S0000276; HSP26.
DR INTERPRO, IPR002068; -
DR PFAM, PF00011; HSP20; 1.
DR PROSITE, PS01031; HSP20; 1.
KW Heat shock.
FT INIT MET 0 0 G -> A (IN REF. 2).
FT CONFLICT 31 31 S -> C (IN REF. 2).
FT CONFLICT 206 206
SEQUENCE 213 AA; 23748 MW; 1C9C4BD0626B6A4 CRC64;

Query Match 17.2%; Score 126; DB 1; Length 213;

Best Local Similarity 29.7%; Pred. No. 0.00011;

Matches 41; Conservative 21; Mismatches 60; Indels 16; Gaps 5;

QY 17 FSEIFAFAFPFAGLPFDRLRLMEDEMEGREYEAELPGVDPDRIHWYRD--GOL 74
ID HS12_ARYSA STANDARD; PRT; 154 AA.
AC P31673;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-JUL-1993 (Rel. 26, Last annotation update)
DT 17.4 KDA CLASS I HEAT SHOCK PROTEIN.
Oryza sativa (Rice).
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.
RN SEQUENCE FROM N.A.
RP [1]
RA Nishi R, Hashimoto H, Uchimiya H, Kato A;
RU Submitted (JUL-1992) to the EMBL/GenBank/DBJ databases.
RL -1- SUBUNIT: FORMS OLIGOMERIC STRUCTURES.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SIMILARITY: BELONGS TO THE SMALL HEAT SHOCK PROTEIN (HSP20)
CC FAMILY.
CC -1- SIMILARITY: PLANTS HAVE FOUR CLASSES OF HSP20: CLASS I AND II
CC WHICH ARE CYTOPLASMIC; CLASS III WHICH IS IN THE CHLOROPLAST
CC AND CLASS IV WHICH IS IN THE ENDOMEMBRANE. THIS PROTEIN BELONGS
CC TO CLASS I.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL, D12635; BAA02160.1; -

DR PIR, JS0710; JS0710.
DR INTERPRO, IPR002068; -
DR PFAM, PF00011; HSP20; 1.
DR PROSITE, PS01031; HSP20; 1.
KW Heat shock; Multigene family.
SQ SEQUENCE 154 AA; 17376 MW; B6A3C4C768598012 CRC64;

Query Match 17.0%; Score 124.5; DB 1; Length 154;
Best Local Similarity 33.6%; Pred. No. 0.0001;

Matches 45; Conservative 20; Mismatches 50; Indels 19; Gaps 6;

QY 12 SLPEFSEIFAFAFPFAGLPFDRLRLMEDEMEGREYEAELPGVDPDRIHWYRD 71
ID HS12_ARYTH STANDARD; PRT; 156 AA.
AC P13853;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DT 17.6 KDA CLASS I HEAT SHOCK PROTEIN (HSP 17.6).
GN HSP17.6.
OS Arabidopsis thaliana (mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, COLUMBIA; TISSUE=LEAF;
RX MEDLINE=90016890; PubMed=2798141;
RA Helm K.W., Vierling E.;
RU "An Arabidopsis thaliana cDNA clone encoding a low molecular weight
RL heat shock protein.";
RL Nucleic Acids Res. 17:7995-7995(1989).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SIMILARITY: BELONGS TO THE SMALL HEAT SHOCK PROTEIN (HSP20)
CC FAMILY.
CC -1- SIMILARITY: PLANTS HAVE FOUR CLASSES OF HSP20: CLASS I AND II
CC WHICH ARE CYTOPLASMIC; CLASS III WHICH IS IN THE CHLOROPLAST
CC AND CLASS IV WHICH IS IN THE ENDOMEMBRANE. THIS PROTEIN BELONGS
CC TO CLASS I.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL, X16076; CAA34208.1; ALT_SEQ.
DR PIR, S06074; S06074.
DR INTERPRO, IPR002068; -
DR PFAM, PF00011; HSP20; 1.
DR PROSITE, PS01031; HSP20; 1.
KW Heat shock; Multigene family.
SQ SEQUENCE 156 AA; 17547 MW; 57BA4B7F1BD637DC CRC64;

Query Match 17.0%; Score 124; DB 1; Length 156;
Best Local Similarity 32.2%; Pred. No. 0.00012;

[illegible]

```

Query Match          16.9%: Score 123.5; DB 1; Length 161;
Best Local Similarity 38.1%: Pred. No. 0.00013;
Matches 37; Conservative 14; Mismatches 39; Indels 7; Gaps 4;

QY 53 RAELPGVDPDKDVLHMVD-GQLTIKAERT---EOKDLDRSEFVAGSFVRYSLPYGA 107
      :|::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 66 KADLPGLTKKE-EVKVEYEDKNVLQSGERSKENEKENDKMWRVERASGKFMRRFLPENA 124
      :|::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

QY 108 DEDDIKATYDKGILTVSV-AVSECGPTPKHIOIRSTN 143
      :|::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

Db 125 KMEYKATMENGVLTVVYPKAPKKPKOVKSIDISGAN 161

RESULT 15
HS11_ORYSA
ID HS11_ORYSA STANDARD: PRT: 150 AA.
AC P27777;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE 16.9 KDA CLASS I HEAT SHOCK PROTEIN.
OS Oryza sativa (Rice).
OC Eukaryota: Viridiplantae; Embryophyta; Tracheophyta;
OC Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.
RN [1]
RP SEQUENCE FROM N.A.

```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 5, 2001, 14:28:41 ; Search time 67.42 Seconds

(Without alignments)
144.019 Million cell updates/sec

Title: US-09-461-774-8

Perfect score: 1 ATTLPVQRHRSLEPFSESL.....SVAVSEKGRPEKHQIRSTN 143

Sequence:

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 195891 seqs, 67900655 residues

Total number of hits satisfying chosen parameters: 195891

Minimum DB seq length: 0
Maximum DB seq length: 2000000000Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	718	98.2	144	2	F70942
2	165.5	22.6	159	2	G70939
3	154.5	21.1	143	2	T40376
4	149	20.4	146	2	S74956
5	148.5	20.3	188	2	A49942
6	147	20.1	140	2	B69496
7	140.5	19.2	151	2	T05739
8	138.5	18.9	142	2	T46658
9	137.5	18.8	157	2	S04939
10	137	18.7	159	2	S72544
11	136.5	18.7	152	2	S72546
12	135	18.5	160	2	P82582
13	133	18.2	147	2	D72385
14	133	18.2	151	2	S21600
15	133	18.2	182	2	A54366
16	131.5	18.0	150	2	S72545
17	131.5	18.0	160	2	T05740
18	131	17.9	149	2	T04171
19	131	17.9	210	2	T09611
20	130.5	17.9	159	1	CYP279
21	130	17.8	151	1	HHWT17
22	130	17.8	229	2	S58210
23	129	17.6	152	2	S23212
24	129	17.6	153	2	S23529
25	129	17.6	156	2	S71566
26	127.5	17.4	154	2	S24396
27	127.5	17.4	154	2	T44801
28	127.5	17.4	161	2	S00646
29	127	17.4	157	2	T14381

30	127	17.4	160	2	JC4377	low-molecular-weig
31	126.5	17.3	155	2	T06449	probable heat shoc
32	126.5	17.3	161	2	S33566	heat shock protein
33	126	17.2	214	2	S45465	heat shock protein
34	124.5	17.0	154	2	J50710	heat shock protein
35	124.5	17.0	159	2	T04173	heat shock protein
36	124	17.0	157	2	S06074	heat shock protein
37	123.5	16.9	156	2	J00351	heat shock protein
38	123.5	16.9	161	2	J00352	heat shock protein
39	122.5	16.8	156	2	T49264	heat shock protein
40	122.5	16.8	163	2	S71567	small heat-shock p
41	121.5	16.6	150	2	S20874	heat shock protein
42	121	16.6	143	2	A57270	heat shock protein
43	121	16.6	159	2	T03958	heat shock protein
44	120	16.4	159	2	T46833	heat-shock protein
45	119.5	16.3	235	2	T06324	heat shock protein

ALIGNMENTS

```
RESULT 1
F70942
probable hspX protein - Mycobacterium tuberculosis (strain H37RV)
N:Alternate names: 14k antigen, 19k major membrane protein
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: F70942; A42651; A43823
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
R:Connor, R.; Davies, R.; Devlin, K.; Feltham, T.; Gentles, S.; Hamlin, N.; Holtroyd,
R:Jandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
N:ature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno
A:Reference number: A70500; MUID:98295987
A:Accession: F70942
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-144 <COL>
A:Cross-references: GB:AL021899, GB:AL123456, NID:g3242282; PIDN:CAA17245.1; PID:g289
A:Experimental source: strain H37RV
R:Verbon, A.; Hartskeerl, R.A.; Schultema, A.; Kolk, A.H.; Young, D.B.; Lathigra, R.
J. Bacteriol. 174, 1352-1359, 1992
A:Title: The 14,000-molecular-weight antigen of Mycobacterium tuberculosis is related
A:Reference number: A42651; MUID:92138631
A:Accession: A42651
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-144 <VER>
A:Cross-references: GB:S79751, NID:g244561; PIDN:AAE21317.1; PID:g244562
R:Lee, B.Y.; Hefta, S.A.; Brennan, P.J.
Infect. Immun. 60, 2066-2074, 1992
A:Title: Characterization of the major membrane protein of virulent Mycobacterium tub
A:Reference number: A43823; MUID:92225631
A:Accession: A43823
A:Molecule type: protein
A:Residues: 2-144 <LEP>
A:Experimental source: strain Erdman
C:Genetics:
A:Gene: hspX
C:Superfamily: alpha-crystallin
C:Keywords: membrane protein; surface antigen
```

Query Match 98.2% Score 718; DB 2; Length 144;
Best Local Similarity 98.6% Pred. No. 2e-58; Mismatches 2; Indels 0; Gaps 0;
Matches 141; Conservative 0;

```
QY 1 ATTLPVQRHRSLEPFSESLFAFPSPAGLRPTFTRLRLDEMKEGRYEVRALPGVD 60
DB 2 ATTLPVQRHRSLEPFSESLFAFPSPAGLRPTFTRLRLDEMKEGRYEVRALPGVD 61
```


Db 62 QITVEENILKSGSKLREKGGKNNYYVERSSAGFERAIRLPDYVERIKAKPEYKNGVL 121

Qy 122 TVSYVAVSESGKPTKEHIQI 139

Db 122 TIRVPKKEER-KKKVIEV 138

RESULT 9
S04939
heat shock 22K protein - Chlamydomonas reinhardtii
C:Species: Chlamydomonas reinhardtii
C:Date: 28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change 13-Aug-1999
C:Accession: S04939
R:Gittman, B.; Ish-Shalom, D.; Even, D.; Glaczinski, H.; Oltersbach, P.; Ohad, I.; Klopst
Eur. J. Biochem. 182, 539-546, 1989
A:Title: The nuclear-coded chloroplast 22-kDa heat-shock protein of Chlamydomonas. Evid
A:Reference number: S04939; MUID:89325298
A:Accession: S04939
A:Molecule type: mRNA
A:Residues: 1-157 <GRI>
A:Cross-references: EMBL:X15053; NID:g18151; PIDN:CAA33152.1; PID:g18152
C:Superfamily: alpha-crystallin
C:Keywords: chloroplast

	Query Match	18.8%	Score 137.5;	DB 2:	length 157;
	Best Local Similarity	38.8%;	Pred. No.	1.4e-05;	
Matches	Conservative	14;	Mismatches	37;	Indels 9; Gaps 4;
QY	50	KVFAAELEGVPPDDVDHVMNRDGLQTIAERT---	EOKDLDS---	RSEFANGSFVRYSL	103
	:::::	:	:.....:	:	:
Db	58	FELNADAPGMPDP-DVKVEIDEGVLMTGERKLSITTT	EAGSGKVRSRSTAYTSFSRAFLS		116
	:	:	:	:	:
QY	104	PVGADEDDEIKATFYKGILTVSVANSE--GRPTKEHIQI			139
	:	:	:	:	:
bB	117	PENANPGGITAMDKGVLTVPKKREPAKPPEPKRIAV			154
	:	:	:	:	:

RESULT 10
 S72544
 heat shock protein 17.9 - pearl millet
 C:Species: Pennisetum americanum (pearl millet)
 C:Date: 01-May-1998 #sequence_revision 01-May-1998 #text_change 20-Jun-2000
 C:Accession: S72544
 R:Cavan, G.P.; Stoelt, K.; Stevens, M.J.; Howarth, C.J.
 Plant Mol. Biol. 30, 1075b, 1996
 Title: Sequence announcement.
 Reference number: S72544
 A:Accession: S72544
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: mRNA
 A:Residues: 1-159 <C>
 A:Cross-references: EMBL:X94193; NID:g1122316; PIDN:CAA63903.1; PID:g1122317
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, December 1995
 A:Note: the source is designated as Pennisetum glaucum
 C:Genetics:
 A:Gene: hsp17.9
 C:Superfamily: alpha-crystallin
 C:Keywords: heat shock; stress-induced protein

Query Match	18.7%	Score 137	DB 2	Length 159
Best Local Similarity	32.8%	Pred. No. 1.6e-05		
Matches	44	Conservative 21	Mismatches 53	Indels 16
			Gaps 5	
QY	12	SLPEEFSELFAPFPSPGLRPTFLRLRLDEMKEGREYVARLELPGVDPOKDVHTWRD	71	
Db	34	SLPFSFPTSTSETAFAAG-----ARIDMKETPEAHVKADVPALAKKE-EVKVEVED	83	
QY	72	GQ-LLTKAERT-----EOKDDLGRSEFAVGSFVPTVSLVPGADEDDIKATYDGKILTVSA	126	
Db	84	GNVLIQISGERKKEDEEKTVDWHRVERSSGKPMRRRLPEENAKTQDIRSMENGVLTTPV	143	

QY	127	VSE-GKPTKEKHQI	139
Db	144	KEEVKKPEVKSQI	157

```

RESULT 11
S72546
heat shock protein 17.0 - pearl millet
C:Species: Pennisetum americanum (pearl millet)
C:Date: 14-Apr-1998 #sequence_revision 15-May-1998 #text_change 20-Jun-2000
C:Accession: S72546
R:Cavan, G.P.; Skoet, K.; Stevens, M.J.; Howarth, C.J.
Plant Mol. Biol. 30, 1076, 1996
A:Title: Sequence announcement.
A:Reference number: S72545
A:Accession: S72546
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: mRNA
A:Residues: 1-152 <CAV>
A:Cross-references: EMBL:X94191; NID:g1122314; PIDD:CAA63901.1; PTD:g1122315
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, December 1995
A:Note: the source is designated as Pennisetum glaucum
A:Superfamily: alpha-crystallin

```

Query Match	18.7%	Score 136.5	DB 2	Length 152
Best Local Similarity	33.1%	Pred. No. 1.6e-05		
Matches 45	Conservative 23	Mismatches 57	Indels 11	Gaps
QY	13	LFPEFSEIFAAPSPAGLRPTFTDRLM---	RLDEEMKRGREYRAELPGVDPKDIVIMV	69
		: : : : : : : : : : : : :		
Db	17	LMDFDSMFSTRIVSAG-SPSDTFAARFAARIDMKETPEAHVFEADLPGYKKE-EVA	EVV	74
QY	70	RDGO-LTIKAERTEQK---DLGRSEEPAYGFWRTVSLPGADEDIDKATYDKGLTVS		124
		: : : : : : : : : : : : : : : : :		
Db	75	EDGNVLIASIGGRSKEKEDKNDNRHVRVERSSQGFRRRLPLGNAKAVDQVAGLGEVLTVT		134
QY	125	V-AVSEGGKPTKEHTOI		139
		: : : : : : : :		
Db	135	VPKAEKKRPVKALEI		150

RESULT 12

low molecular weight heat shock protein XE2234 [imported] - *Xylella fastidiosa* (strain F82582)

C:Species: *Xylella fastidiosa*

C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000

C:Accession: F82582

R:anonymous, The *Xylella fastidiosa* Consortium of the Organization for Nucleotide Seq

Nature 406, 151-157, 2000

A:Title: The genome sequence of the plant pathogen *Xylella fastidiosa*.

A:Reference number: A82515; WUJID:20365717

A:Note: for a complete list of authors see reference number A59328 below

A:Accession: F82582

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-160 <SIAM>

A:Cross-references: GB:AE004036; GB:AE003849; NID:J9107384; PIDN:AAEF85033.1; GSPDB:CNF85033.1; GSPDB:CNF85033.1

A:Experimental source: strain 9a5c

R:Stimpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; Britones, M.R.S.; Bueno, M.L.P.; Camargo, A.V.; Camargo, L.E.A.; Carraro, D.M.; Carreira, S.; Neto, E.; Docena, C.; El-Dorri, H.; Facchinati, A.P.; Ferreira, A.J.S. submitted to GenBank, June 2000

A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Fradinho, J.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kurame, E.E.; Machado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, A.; Authors: Martins, E.M.F.; Matsushima, A.Y.; Menck, C.F.M.; Mirecia, E.C.; Miyaki, C.; F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawaia, A.; Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silva, T.; Tshaharo, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.

Reference number: A59328

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 5, 2001, 14:27:33 ; Search time 45.19 Seconds
(without alignments)
56.824 Million cell updates/sec

Title: US-09-461-774-8

Perfect score: 731

Sequence: 1 ATTLPVQRHRSLEPFPESEL.....SVAVSCKPTKEKHQIRSTN 143

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 174772 seqs, 17957048 residues

tal number of hits satisfying chosen parameters: 174772

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/PCTUS.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/Backfilest.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	74.5	10.2	199	2	US-08-900-407-4 Sequence 4, Appli
2	71.5	9.8	193	2	US-08-900-407-3 Sequence 3, Appli
3	69.5	9.5	438	1	US-08-261-110A-2 Sequence 2, Appli
4	69	9.4	388	1	US-08-261-110A-4 Sequence 4, Appli
5	68.5	9.4	545	2	US-08-467-822-30 Sequence 30, Appli
6	68.5	9.4	546	3	US-08-470-260-6 Sequence 6, Appli
7	68.5	9.4	546	3	US-08-471-491-6 Sequence 6, Appli
8	68.5	9.4	546	3	US-08-466-662-6 Sequence 6, Appli
9	68	9.3	1528	1	US-08-463-092B-6 Sequence 6, Appli
10	68	9.3	1528	2	US-08-462-109A-6 Sequence 6, Appli
11	68	9.3	1528	2	US-08-460-907B-6 Sequence 6, Appli
12	68	9.3	1528	3	US-08-463-179A-6 Sequence 6, Appli
13	68	9.3	1528	3	US-08-461-384B-6 Sequence 6, Appli
14	67	9.2	539	2	US-08-838-189D-2 Sequence 2, Appli
15	67	9.2	539	2	US-08-852-344D-2 Sequence 2, Appli
16	67	9.2	539	3	US-08-344-639E-2 Sequence 2, Appli
17	67	9.2	539	3	US-08-344-639E-2 Sequence 2, Appli
18	66	9.0	1255	2	US-08-899-595-1 Sequence 1, Appli
19	66	9.0	1255	2	US-08-899-595-1 Sequence 1, Appli
20	65.5	9.0	153	3	US-09-048-889-2 Sequence 2, Appli
21	65.5	9.0	642	2	US-08-818-253-2 Sequence 2, Appli
22	65.5	9.0	642	2	US-08-818-253-6 Sequence 6, Appli
23	65.5	9.0	656	2	US-08-818-253-4 Sequence 4, Appli
24	65.5	9.0	656	2	US-08-818-253-8 Sequence 8, Appli
25	65.5	9.0	797	3	US-08-433-522A-2 Sequence 2, Appli
26	65.5	9.0	797	3	US-08-433-522A-4 Sequence 4, Appli
27	65.5	9.0	797	3	US-08-433-522A-6 Sequence 6, Appli
28	65.5	9.0	797	3	US-09-135-166-2 Sequence 2, Appli

29	65.5	9.0	797	3	US-09-135-166-4 Sequence 4, Appli
30	65.5	9.0	797	3	US-09-135-166-6 Sequence 6, Appli
31	65	8.9	969	4	PCT-US93-03027-4 Sequence 4, Appli
32	65	8.9	1093	4	PCT-US93-03077-1 Sequence 1, Appli
33	64.5	8.8	793	3	US-08-433-522A-10 Sequence 10, Appli
34	64.5	8.8	793	3	US-09-135-166-10 Sequence 10, Appli
35	64	8.8	969	1	US-08-365-689-3 Sequence 3, Appli
36	64	8.8	969	1	US-07-747-781-3 Sequence 3, Appli
37	64	8.8	969	1	US-08-145-138A-3 Sequence 3, Appli
38	64	8.8	969	4	PCT-US92-06888-3 Sequence 3, Appli
39	63.5	8.7	997	1	US-08-324-977-50 Sequence 50, Appli
40	63.5	8.7	997	2	US-08-384-616-50 Sequence 50, Appli
41	63.5	8.7	997	2	US-08-904-686A-50 Sequence 50, Appli
42	63.5	8.7	1349	2	US-08-612-734B-2 Sequence 2, Appli
43	63.5	8.7	2620	2	US-08-324-977-32 Sequence 32, Appli
44	63.5	8.7	2620	2	US-08-384-616-32 Sequence 32, Appli
45	63.5	8.7	2620	2	US-08-904-686A-32 Sequence 32, Appli

ALIGNMENTS

RESULT 1
US-08-900-407-4
Sequence 4, Application US/08900407
Patent No. 5962262
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer
APPLICANT: Lal, Preeti
APPLICANT: Corley, Neil C.
APPLICANT: Shah, Purvi
TITLE OF INVENTION: NEW HUMAN HEAT SHOCK 27 LIKE
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESSES:
ADDRESS: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Dr.
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/900.407
FILING DATE: Filed Herewith
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0351 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-845-4166
TELEFAX: 415-845-0555
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 199 amino acids
TYPE: amino acid
STRANDEDNESS: Single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 662841
US-08-900-407-4
Query Match 10.2%; Score 74.5; DB 2; Length 199;
Best Local Similarity 25.0%; Pred. No. 0.24;

Matches	32;	Conservative	21;	Mismatches	64;	Indels	11;	Gaps	5
OY	22	AAFPSPAGLRPTFFDRLMR-LEDEKKE----	GREYVRAELPGVDDPKVHIMVRDQQLT	75					
Db	61	AAIESPPAAAPAYSRALSRQLSSGSEIRHTADRMVESLDVNHFAFD-ELTVKTKGAYE	119						
OY	76	IAAEETEEKDDLGRSEFAFGSFVRYSLPVGADDEDDIKATYD-KGLITYSVAVSEBKPTPE	134						
Db	120	ITGKHEERODEENG--YISRCTFRKRYTLPGGVDPPQVSSSLSPGEGTLTVAEAMPKLATOS	176						
OY	135	KHIOIRST 142							
Db	177	NEITIPVT 184							

```

RESULT 2
; Sequence 3, Application US/08900407
; Patent No. 5962262
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer
; APPLICANT: Lal, Preeti
; APPLICANT: Corley, Neil C.
; APPLICANT: Shah, Purvi
; TITLE OF INVENTION: NEW HUMAN HEAT SHOCK 27 LIKE
; TITLE OF INVENTION: PROTEIN
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Dr.
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/900,407
; FILING DATE: Filed Herewith
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0351 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 193 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 63522
; US-08-900-407-3

```

[illegible]

```

RESULT      3
US-08-261-110A-2
Sequence 2, Application US/08261110A
Patent No. 5674992
GENERAL INFORMATION:
APPLICANT: JACENDORF, ANDRE
APPLICANT: CERUTTI, HERIBERTO
TITLE OF INVENTION: CDNA ENCODING A RECA HOMOLOG IN
TITLE OF INVENTION: EUKARYOTES
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON, HARGRAVE, DEVANS & DOYLE LLP
STREET: CLINTON SQUARE, P.O. BOX 1051
CITY: ROCHESTER
STATE: NEW YORK
COUNTRY: USA
ZIP: 14603
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/261,110A
FILING DATE: 16-JUN-1994
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/938,332
FILING DATE: 28-AUG-1992
ATTORNEY/AGENT INFORMATION:
NAME: TIMIAN, SUSAN J.
REGISTRATION NUMBER: 34,103
REFERENCE/DOCKET NUMBER: 19603/10231(D-1292A)
TELECOMMUNICATION INFORMATION:
TELEPHONE: (716) 263-1636
TELEFAX: (716) 263-1600
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 438 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-261-110A-2

Query Match      9.5%; Score 69.5; DB 1; Length 438;
Best Local Similarity 23.9%; Pred. No. 3.1;
Matches 34; Conservative 22; Mismatches 59; Indels 27; Gaps 6

QY      13 LPEFESELFNAAP-----SFAGLRPTFDRMLRLDEDM-----REGREYVRELPP----- 57
Dp      10 LNSFETPLSELPFPTCCSFSSPLRFSSCYSRRLTSPVTYAAKKLSHKLSSEEDDRING 69
QY      58 GVPDDKDVHIMVWDGOLTTIKARTEQKLDGRSEFAYGSFVRTVSLPVGADDEDIDIKATVD 117
Dp      70 ALSPDADSRLDQ-----KALEAMNNDIN--SSRGKGSYTR-----LGSAGCALVETFS 117
QY      118 KGLITVSAVASEGKPTKEHIQT 139
Dp      118 SGILTLDLALGGGLPKGRVVEI 139

RESULT      4
US-08-261-110A-4

```

```
; Sequence 4, Application US/08261110A
; Patent No. 5674992
; GENERAL INFORMATION:
; APPLICANT: JAGENDORF, ANDRE
; APPLICANT: CEROTTI, HERIBERTO
; TITLE OF INVENTION: CDNA ENCODING A REGA HOMOLOG IN
; TITLE OF INVENTION: EUKARYOTES
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON, HARGRAVE, DEVANS & DOYLE LLP
; STREET: CLINTON SQUARE, P.O. BOX 1051
; CITY: ROCHESTER
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/261.110A
; FILING DATE: 16-JUN-1994
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/938,332
; FILING DATE: 28-AUG-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: TIMIAN, SUSAN J.
; REGISTRATION NUMBER: 34,103
; REFERENCE/DOCKET NUMBER: 19603/10231(D-1292A)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1636
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 388 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-261-110A-4

Query Match          9.4%; Score 69; DB 1; Length 388;
Best Local Similarity 27.2%; Pred. No. 3;
Matches 22; Conservative 14; Mismatches 33; Indels 12; Gaps 3;

59 VDPDDVHIMVRDGLTIKAERTEQKDLGRSEFAVGSFRTVSLPVGADDDIKATYDK 118
: |||: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 21 LSPDADSRRLDQ-----KALEAMNDIN--SFGKGSYTR-----LGSAGCALVETFS 68

Oy 119 GILTVSAVASEGKPTKEKHIQI 139
|||: | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | |
Db 69 GILTLDALGGILPKGRVVEI 89

RESULT 5
US-08-467-822-30
; Sequence 30, Application US/08467822
; Patent No. 5843460
; GENERAL INFORMATION:
; APPLICANT: Labigne, Agnes
; APPLICANT: Sauerbaum, Sebastien
; APPLICANT: Ferrero, Richard L.
; APPLICANT: Thibierge, Jean-Michel
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS AGAINST
; TITLE OF INVENTION: HELICOBACTER INFECTION. POLYPEPTIDES FOR USE IN THE
; TITLE OF INVENTION: COMPOSITIONS, AND NUCLEIC ACID SEQUENCES ENCODING SAID
; TITLE OF INVENTION: POLYPEPTIDES
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flinnegan, Henderson, Farabow, Garrett &
```

```
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/467,822
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/447,177
; FILING DATE: 19-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/432,697
; FILING DATE: 02-MAY-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Meyers, Kenneth J.
; REGISTRATION NUMBER: 25,146
; REFERENCE/DOCKET NUMBER: 03495.0137-02000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 408-4400
; TELEFAX: (202) 408-4400
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 545 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: Protein
; LOCATION: 1..545
; OTHER INFORMATION: /product= "H. pylori - Hsp B."
; US-08-467-822-30

Query Match          9.4%; Score 68.5; DB 2; Length 545;
Best Local Similarity 25.5%; Pred. No. 5.8;
Matches 26; Conservative 16; Mismatches 53; Indels 7; Gaps 2;

Oy 42 EDEMEGRYEVRAELPGVDPDKVHIMVRDGLTIKAERTEQKDLGRSEFAVGSFRTV 101
| |||: | | | | | | | | | | | | | | | | | | | | | |
: | | | | | | | | | | | | | | | | | | | | | |
: | | | | | | | | | | | | | | | | | | | | | |
Db 384 EYEMEKRRVVDALSATAAVEGIVIGGALIRAAOKVHLNLDHDEKVGVEIIMRAI 443

Oy 102 SLPGADEDDIKATYDKGILTVSAVASEGKPTKEKHIQIRSTN 143
| | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | |
Db 444 KAPLA--QIAINAGYDGG-----VYVNEVEKHEGHEGFNASN 478

RESULT 6
US-08-470-260-6
; Sequence 6, Application US/08470260
; Patent No. 6077706
; GENERAL INFORMATION:
; APPLICANT: Covacci, Antonello
; APPLICANT: Bugnoli, Massimo
; APPLICANT: Telford, John
; APPLICANT: Macchia, Giovanni
; APPLICANT: Rappuoli, Rino
; TITLE OF INVENTION: Helicobacter pylori Proteins Useful
; TITLE OF INVENTION: for Vaccines and Diagnostics
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
```

```

? CITY: Emeryville
? STATE: California
? COUNTRY: USA
? ZIP: 94608-2916
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patentln Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/470,260
? FILING DATE:
? CLASSIFICATION:
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/256,848
? FILING DATE: 21-OCT-1994
? ATTORNEY/AGENT INFORMATION:
? NAME: McClung, Barbara G.
? REGISTRATION NUMBER: 33,113
? REFERENCE/DOCKET NUMBER: 0316.001
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (510) 601-2708
? TELEFAX: (510) 655-3542
? INFORMATION FOR SEQ ID NO: 6:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 546 amino acids
? TYPE: amino acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: protein
?
US-08-470-260-6

```

```

Query Match          9.4%; Score 68.5; DB 3; Length 546;
Best Local Similarity 25.5%; Pred. No. 5.8;
Matches 26; Conservative 16; Mismatches 53; Indels 7; Gaps 2;

QY 42 EDMKRGRYVRAELPGVDPDKVHIMVRDGLTIKAERTQKDLGRSEFAYGSFVRV 101
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 385 EVMKEKKRDVDALSTKAAVEGIVIGGALLIRAAQKVHNLHDEKVGTEIIMRAI 444
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 102 SLPGADEDDIKATYDKGILTVSAVASEGKPTKEKHIOIRSTN 143
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 445 KAPLA--QIAINAGYDGG-----VVVNEVEKHEHGFGFNASN 479
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

RESULT 7
US-08-471-491-6
; Sequence 6, Application US/08471491B
; Patent No. 6090611
; GENERAL INFORMATION:
; APPLICANT: Covacci, Antonello
; APPLICANT: Bugnoli, Massimo
; APPLICANT: Telford, John
; APPLICANT: Macchia, Giovanni
; APPLICANT: Rappuoli, Rino
; TITLE OF INVENTION: Helicobacter Pylori Proteins Useful For Vaccines And
; TITLE OF INVENTION: Diagnostics
; FILE REFERENCE: CHIR0044
; CURRENT APPLICATION NUMBER: US/08/471,491B
; CURRENT FILING DATE: 1995-06-06
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 6
; LENGTH: 546
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-08-471-491-6

```

```

Query Match          9.4%; Score 68.5; DB 3; Length 546;
Best Local Similarity 25.5%; Pred. No. 5.8;
Matches 26; Conservative 16; Mismatches 53; Indels 7; Gaps 2;

```

```

QY 42 EDMKRGRYVRAELPGVDPDKVHIMVRDGLTIKAERTQKDLGRSEFAYGSFVRV 101
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 385 EVMKEKKRDVDALSTKAAVEGIVIGGALLIRAAQKVHNLHDEKVGTEIIMRAI 444
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 102 SLPGADEDDIKATYDKGILTVSAVASEGKPTKEKHIOIRSTN 143
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 445 KAPLA--QIAINAGYDGG-----VVVNEVEKHEHGFGFNASN 479
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

Query Match          9.4%; Score 68.5; DB 3; Length 546;
Best Local Similarity 25.5%; Pred. No. 5.8;
Matches 26; Conservative 16; Mismatches 53; Indels 7; Gaps 2;

QY 42 EDMKRGRYVRAELPGVDPDKVHIMVRDGLTIKAERTQKDLGRSEFAYGSFVRV 101
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 385 EVMKEKKRDVDALSTKAAVEGIVIGGALLIRAAQKVHNLHDEKVGTEIIMRAI 444
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 102 SLPGADEDDIKATYDKGILTVSAVASEGKPTKEKHIOIRSTN 143
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 445 KAPLA--QIAINAGYDGG-----VVVNEVEKHEHGFGFNASN 479
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

RESULT 9
US-08-463-092B-6
; Sequence 6, Application US/08463092B
; Patent No. 5766880
; GENERAL INFORMATION:
; APPLICANT: Cole, Susan P.C.
; APPLICANT: Deeley, Roger G.
; TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULES ENCODING
; TITLE OF INVENTION: MULTIDRUG RESISTANCE PROTEINS
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PARTEQ RESEARCH & DEVELOPMENT INNOVATIONS
; STREET: Queen's University at Kingston
; CITY: Kingston
; STATE: Ontario
; COUNTRY: CANADA
; ZIP: K7L 3N6
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/463,092B
; FILING DATE: 05-JUN-1995
; CLASSIFICATION: 435

```

```

      FILING DATE: 27-OCT-1992
      APPLICATION NUMBER: 08/029,340
      FILING DATE: 8-MAR-1993
      APPLICATION NUMBER: 08/141,893
      FILING DATE: 26-OCT-1993
      APPLICATION NUMBER: 08/407,207
      FILING DATE: 20-MAR-1995
      ATTORNEY/AGENT INFORMATION:
        NAME: DeConlt, Giulio A. Jr.
        REGISTRATION NUMBER: 31,503
        REFERENCE/DOCKET NUMBER: P01-002CP4
        TELECOMMUNICATION INFORMATION:
          TELEPHONE: (617) 227-7400
          TELEFAX: (617) 227-5941
        INFORMATION FOR SEQ ID NO: 6:
          SEQUENCE CHARACTERISTICS:
            LENGTH: 1528 amino acids
            TYPE: amino acid
            TOPOLOGY: linear
        MOLECULE TYPE: protein
      US-08-462-109A-6

Query Match           9.3% Score 68; DB 2; Length 1528;
Best Local Similarity 28.0%; Pred. NO. 31;
Matches 26; Conservative 14; Mismatches 31; Indels 22; Gaps 4

QY   61 PDKVIVHWRGGLTIKAKERTEQKLDCRSEFAVGSFVRTVSLPGAGDEDDIKATVDKGI 120
      | ||||| : ||: | || | || | : || | : || | : |
Db    633 PQYDVIIVMSGKISSEMSTYQLDRDG---AFAEFLRTYA-----MEQDLASEDD---- 881
QY   121 LTVSAVASEGKPTPE-----KHQIRSTN 143
      : || : | || | || | || | : |
Db    882 -SVSGSGEKSRPVENGMLVTDTVGKHLQRHLSN 913

RESULT 11
US-08-460-907B-6
; Sequence 6, Application US/08460907B
; Patent No. 5891724
; GENERAL INFORMATION:
; APPLICANT: Deeley, Roger G.
; APPLICANT: Cole, Susan P.C.
; TITLE OF INVENTION: METHODS FOR CONFERRING MULTIDRUG
; TITLE OF INVENTION: RESISTANCE ON A CELL
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PARTS RESEARCH & DEVELOPMENT INNOVATIONS
; STREET: Queen's University at Kingston
; CITY: Kingston
; STATE: Ontario
; COUNTRY: CANADA
; ZIP: K7L 3N6
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/460,907B
; FILING DATE: 05-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/966,923
; FILING DATE: 27-OCT-1992
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/029,340
; FILING DATE: 8-MAR-1993
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/141,893
; FILING DATE: 26-OCT-1993

```


[illegible]

```

QY      68 MYRSQLITKAKERTQOKLDSGRSEFVAGSFRTVSLPGADDEDDIKATYDKGILTVSVAV 127
          | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db      93 -----IVTNQESNENTDPTRRSFCGVIGTIALGVATSAQ-----TTAAVAL 134

QY      128 SEGKPEKHIO-----IRSTN 143
          | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db      135 VEAKAKSDIEKKAIRDTN 155

RESULT 15
US-08-838-189D-2
: Sequence 2, Application US/08838189D
: Patent No. 5998169
: GENERAL INFORMATION:
:   APPLICANT: KLEIN, Michel H
:   APPLICANT: DU, Run-Pan
:   APPLICANT: EMASYSHN, Mary E
:   TITLE OF INVENTION: MULTIMERIC HYBRID GENE ENCODING A
:   TITLE OF INVENTION: CHIMERIC PROTEIN WHICH CONFERS PROTECTION AGAINST
:   TITLE OF INVENTION: PARAINFLUENZA VIRUS AND RESPIRATORY SYNCYTIAL VIRUS
:   NUMBER OF SEQUENCES: 38
:   CORRESPONDENCE ADDRESS:
:     ADDRESSEE: Sim & Mcburney
:     STREET: 6th Floor, 330 University Avenue
:     CITY: Toronto
:     STATE: Ontario
:     COUNTRY: Canada
:     ZIP: M5G 1R7
: COMPUTER READABLE FORM:
:   MEDIUM TYPE: Floppy disk
:   COMPUTER: IBM PC compatible
:   OPERATING SYSTEM: PC-DOS/MS-DOS
:   SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
:   APPLICATION NUMBER: US/08/838,189D
:   FILING DATE: 16-APR-1997
:   CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
:   APPLICATION NUMBER: US 08/001,554
:   FILING DATE: 06-JAN-1993
: PRIOR APPLICATION DATA:
:   APPLICATION NUMBER: GB 9200117.1
:   FILING DATE: 06-JAN-1992
: ATTORNEY/AGENT INFORMATION:
:   NAME: STEWART, Michael I
:   REGISTRATION NUMBER: 24,973
: REFERENCE/DOCKET NUMBER: 1038-687 MIS:jib
: TELECOMMUNICATION INFORMATION:
:   TELEPHONE: (416) 595-1155
:   TELEFAX: (416) 595-1163
: INFORMATION FOR SEQ ID NO: 2:
:   SEQUENCE CHARACTERISTICS:
:     LENGTH: 539 amino acids
:     TYPE: amino acid
:     STRANDEDNESS: single
:     TOPOLOGY: linear
: US-08-838-189D-2

Query Match      9.2%: Score 67; DB 2; Length 539;
Best Local Similarity 21.3%; Pred. No. 8.6;
Matches 30; Conservative 23; Mismatches 46; Indels 42; Gaps 6.

QY      25 PSFAGLRPFDFRLM-----RLEDMMKRGREV-----RAELP---GVDPKDVIH 67
          | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db      35 PFGMKISONEFERIYLILSLIPKIEDNSCGGDQOIROYKRLDLRIIPLYDGRLQKDV-- 92

QY      68 MYRSQLITKAKERTQOKLDSGRSEFVAGSFRTVSLPGADEDDIKATYDKGILTVSVAV 127
          | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db      93 -----IVTNQESNENTDPTRRSFCGVIGTIALGVATSAQ-----TTAAVAL 134

QY      128 SEGKPEKHIO-----IRSTN 143
          | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :

```

Tue Jan 16 10:17:32 2001

us-09-461-774-8.ra1

Page 8

Db 135 VEAQAKSDIEKLEAIRDTN 155

Search completed: January 5, 2001, 14:27:36
Job time: 228 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 5, 2001, 14:26:43 ; Search time 62.98 Seconds
(without alignments)
77.639 Million cell updates/sec

Title: US-09-461-774-8

Perfect score: 731
Sequence: 1 ATTLPGVGRHRSLEPERSL.....SVAVSECKPREKHQIRSTN 143

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 268485 seqs, 34193795 residues
Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08
Maximum Match 1008
Listing first 45 summaries

Database :

A.Geneseq.36:*

- 1: /SIDSL/gcgdata/geneseq/geneseq/AA1980.DAT:*
- 2: /SIDSL/gcgdata/geneseq/geneseq/AA1981.DAT:*
- 3: /SIDSL/gcgdata/geneseq/geneseq/AA1982.DAT:*
- 4: /SIDSL/gcgdata/geneseq/geneseq/AA1983.DAT:*
- 5: /SIDSL/gcgdata/geneseq/geneseq/AA1984.DAT:*
- 6: /SIDSL/gcgdata/geneseq/geneseq/AA1985.DAT:*
- 7: /SIDSL/gcgdata/geneseq/geneseq/AA1986.DAT:*
- 8: /SIDSL/gcgdata/geneseq/geneseq/AA1987.DAT:*
- 9: /SIDSL/gcgdata/geneseq/geneseq/AA1988.DAT:*
- 10: /SIDSL/gcgdata/geneseq/geneseq/AA1989.DAT:*
- 11: /SIDSL/gcgdata/geneseq/geneseq/AA1990.DAT:*
- 12: /SIDSL/gcgdata/geneseq/geneseq/AA1991.DAT:*
- 13: /SIDSL/gcgdata/geneseq/geneseq/AA1992.DAT:*
- 14: /SIDSL/gcgdata/geneseq/geneseq/AA1993.DAT:*
- 15: /SIDSL/gcgdata/geneseq/geneseq/AA1994.DAT:*
- 16: /SIDSL/gcgdata/geneseq/geneseq/AA1995.DAT:*
- 17: /SIDSL/gcgdata/geneseq/geneseq/AA1996.DAT:*
- 18: /SIDSL/gcgdata/geneseq/geneseq/AA1997.DAT:*
- 19: /SIDSL/gcgdata/geneseq/geneseq/AA1998.DAT:*
- 20: /SIDSL/gcgdata/geneseq/geneseq/AA1999.DAT:*
- 21: /SIDSL/gcgdata/geneseq/geneseq/AA2000.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	718	98.2	144	20	W94900
2	718	98.2	331	20	Y32060
3	718	98.2	368	20	Y32069
4	718	98.2	433	20	Y32065
5	718	98.2	856	20	Y32064
6	356	48.7	131	12	R12489
7	238	32.6	48	16	R85672
8	238	32.6	48	17	R92889
9	238	32.6	48	18	W18183
10	238	32.6	48	18	W75700
11	113.5	15.5	181	14	R43654
12	111	15.2	153	6	P50065

13	104.5	14.3	154	6	P50066
14	101.5	13.9	148	10	P90455
15	101.5	13.9	253	16	R79057
16	94	12.9	143	6	P50063
17	92	12.6	149	6	P50064
18	85	11.6	460	20	Y29294
19	82	11.2	918	20	Y35654
20	81.5	11.1	156	21	Y77926
21	79	10.8	175	20	Y07074
22	77	10.5	175	17	R87512
23	76	10.4	975	19	W72746
24	75.5	10.3	1528	21	W20056
25	75.5	10.3	1552	21	R85577
26	75.5	10.3	1583	18	W20057
27	74.5	10.2	205	14	R41022
28	74.5	10.2	471	19	W54352
29	72	9.8	405	20	Y34438
30	72	9.8	461	20	Y34438
31	69.5	9.5	438	18	W34668
32	69.5	9.5	472	21	W84443
33	69.5	9.5	795	20	W97842
34	69	9.4	283	20	Y37057
35	68.5	9.4	545	16	R74339
36	68.5	9.4	545	16	R67374
37	68.5	9.4	545	17	W06732
38	68.5	9.4	546	14	R41661
39	68.5	9.4	546	14	R41200
40	68.5	9.4	546	19	W98776
41	68	9.3	423	20	Y37283
42	68	9.3	1528	19	W57487
43	68	9.3	1528	20	W99895
44	68	9.3	1528	20	W74472
45	68	9.3	1528	21	Y55800

ALIGNMENTS

RESULT 1	
ID W94900	standard; Protein: 144 AA.
XX	
AC W94900;	
XX	
DT 11-MAY-1999	(first entry)
XX	
DE Alpha-crystallin heat shock protein.	
XX	
KW Tuberculosis; Mycobacterium; alpha-crystallin heat shock protein;	
KW acr gene; attenuated; vaccine; antimycobacterial; pathogenic.	
XX	
OS Mycobacterium sp.	
XX	
PN W09902670-A1.	
XX	
PD 21-JAN-1999.	
XX	
PF 09-JUL-1998;	98W0-US14227.
XX	
PR 10-JUL-1997;	97US-0052199.
XX	
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.	
XX	
PI Barry CE, Crane D, Yuan Y;	
XX	
DR WPI: 1999-120871/10.	
XX	
DR N-PSDB; X11761.	
XX	
PT Novel attenuated strains of Mycobacterium tuberculosis and M. bovis	
PT - useful in the production of a vaccine against tuberculosis,	
PT achieved by reducing or eliminating expression of the	
PT alpha-crystallin heat shock protein gene	
XX	

Soybean heat shock
Antigens of Mycob
Glycosylphosphatid
Soybean heat shock
Soybean heat shock
Protein encoded by
A. thaliana enviro
Renal cancer assoc
Human alpha B crys
Drosophila kinesin
C. elegans UNC-53
C. elegans UNC-53
Protein homologous
Heat shock 27 kD p
Porphyromonas ging
Porphyromonas ging
Arabidopsis thalia
Amino acid sequenc
Human P2Y11 recept
Protein which is s
Helicobacter pylori
Heat shock protein
H. pylori heat sho
H. pylori hsp60.
Hsp. Helicobacter
H. pylori GPO 118
Amino acid sequenc
Murine multidrug r
Mouse multidrug re
Mouse multidrug re
Murine multidrug re

PS Examples: Page 25; 33pp; English.

CC The invention relates to a vaccine for protection against tuberculosis in
CC which Mycobacterium sp. are attenuated by having the expression of the
CC alpha-crystallin heat shock protein gene (acr gene) reduced by at least
CC 75 percent. The attenuated strains can be used as models for screening
CC for novel antimycobacterial agents. Also, the mixture of Mycobacterium
CC sp. cellular debris can be used for detection of exposure to M. bovis
CC and M. tuberculosis, used for commercial adjuvants, used as diagnostics
CC including use as reagents for the development of monoclonal antibodies
CC to recognize wild type mycobacteria in patient samples. The vaccine is
CC based on M. tuberculosis and is obtained without the multitude of
CC passages so that the immune response generated is more relevant to the
CC wild type, pathogenic bacteria. The present sequence represents the
CC alpha-crystallin heat shock protein.

XX Sequence 144 AA:

Query Match 98.2%; Score 718; DB 20; Length 144;
Best Local Similarity 98.6%; Pred. No. 2; 1e-73;
Matches 141; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATTLPVGRHRSLEPSESELPAAFPSPAGLRPTDTRMLRLEDEMKRGREVRNELPGVD 60
DB 2 attlpvgqhprslfpetselfaafpsfaglrptdtrlmrledemkegyevraelpgvd 61
QY 61 PDKDVHIMVRDQGLTTIAERTQKDLGRSEFAYGSEFVRVSLPVGADDDIKATYKGI 120
DB 62 pdkdvdlmvrddqglttlaerteqkdlgrsefaygstvrvtslpvgadeddikatykgi 121
QY 121 LTVSAVASEGKPTKEKHQIRSTN 143
DB 122 ltvsavasegkptekhqrstn 144

RESULT 2
ID Y32060 standard; Protein; 331 AA.

XX AC X32060;

DT 17-JAN-2000 (first entry)

DE Mycobacterium tuberculosis antigen fusion protein MtB39A.

KW Tuberculosis; antigen; fusion protein; MtB39A; ERD14; DPV; MTI;
diagnosis; therapy; vaccine; immunogen.

XX Mycobacterium tuberculosis.

XX Key Location/Qualifiers
FH Peptide 1..7
FT /note= "Met/His tag"
FT 8..151
FT /note= "ERD14"
FT Protein 154..235
FT /note= "DPV"
FT 238..331
FT /note= "MTI"
XX
XX MO9951748-A2.

PN 14-OCT-1999.

PD 07-APR-1999; 99MO-US07717.

XX 07-APR-1998; 9805-0056556.

PR 30-DEC-1998; 98US-0223040.

XX (CORI-) CORIXA CORP.

PA Skelky YAW, Alderson M, Campos-Neto A;

PI

XX WPI; 1999-601610/51.
DR N-PSDB; 220195.
XX
PT New fusion proteins useful for diagnosis, prevention and treatment of
PT tuberculosis -
XX
PS Claim 1; Fig 2; 83pp; English.

CC This sequence represents a recombinant Mycobacterium tuberculosis
CC rti-antigen fusion protein, termed MtB39A, composed of the antigens
CC ERD14, DPV and MTI. The fusion protein is expressed in host cells
CC using a vector carrying a polynucleotide (see 220195) comprising
CC the 3 coding sequences for the antigens. The invention provides
CC fusion proteins (see Y32059-71) containing at least 2 M.
CC tuberculosis antigens. The new fusion proteins and polynucleotides
CC encoding them are useful as vaccines for preventing tuberculosis
CC (claimed), for diagnosis (via in vitro assays or intradermal skin
CC tests for detection of anti-M. tuberculosis antibodies), monitoring
CC of disease progression, and treatment of tuberculosis. They are
CC more effective immunogens than mixtures of the individual protein
CC components.

SQ Sequence 331 AA:

Query Match 98.2%; Score 718; DB 20; Length 331;
Best Local Similarity 98.6%; Pred. No. 7e-73;
Matches 141; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATTLPVGRHRSLEPSESELPAAFPSPAGLRPTDTRMLRLEDEMKRGREVRNELPGVD 60
DB 9 attlpvgqhprslfpetselfaafpsfaglrptdtrlmrledemkegyevraelpgvd 68
QY 61 PDKDVHIMVRDQGLTTIAERTQKDLGRSEFAYGSEFVRVSLPVGADDDIKATYKGI 120
DB 69 pdkdvdlmvrddqglttlaerteqkdlgrsefaygstvrvtslpvgadeddikatykgi 128
QY 121 LTVSAVASEGKPTKEKHQIRSTN 143
DB 129 ltvsavasegkptekhqrstn 151

RESULT 3

ID Y32069 standard; Protein; 368 AA.

XX AC Y32069;

DT 17-JAN-2000 (first entry)

DE Mycobacterium tuberculosis antigen fusion protein MtB36f.

KW Tuberculosis; antigen; fusion protein; MtB36f; ERD14; DPV; MTI;
diagnosis; therapy; vaccine; immunogen.

XX Mycobacterium tuberculosis.

XX Key Location/Qualifiers
FH Peptide 1..8
FT /note= "Met/His tag"
FT 9..151
FT /note= "ERD14"
FT Protein 154..235
FT /note= "DPV"
FT 238..368
FT /note= "MTI"
XX
XX MO9951748-A2.

PN 14-OCT-1999.

PD 07-APR-1999; 99MO-US07717.

XX 07-APR-1998; 9805-0056556.

PR 30-DEC-1998; 98US-0223040.

XX (CORI-) CORIXA CORP.

PA Skelky YAW, Alderson M, Campos-Neto A;

PI

	/note= "MTI"
FT	/note= "MSL"
FT	Protein
PN	MO9951748-A2.
XX	
PD	14-OCT-1999.
XX	
PF	07-APR-1999; 99WO-US07717.
XX	
PR	07-APR-1998; 98US-0056556.
PR	30-DEC-1998; 98US-0223040.
XX	
PA	(CORI-) CORIXA CORP.
XX	
PI	Skeiky YAM, Alderson M, Campos-Neto A;
XX	
DR	WPI; 1999-601610/51.
DR	N-Psdb; 220200.
XX	
PT	New fusion proteins useful for diagnosis, prevention and treatment of tuberculosis -
PS	
XX	
PS	Claim 1; Fig 7A-B; 83pp; English.
XX	
CC	This sequence represents a recombinant Mycobacterium tuberculosis ^T -
CC	tetra-antigen fusion protein, termed Mb46f, composed of the antigenic
CC	ERH1A, DPV, MTI and MSU. The fusion protein is expressed in
CC	host cells using a vector carrying a polynucleotide (see Z20200)
CC	providing the coding sequences for the 4 antigens. The invention
CC	comprises fusion proteins (see Y32059-71) containing at least 2 M.

CC claimed, for diagnosis (via in vitro assays or intradermal skin
CC tests for detection of anti-*M. tuberculosis* antibodies), monitoring
CC tests for disease progression, and treatment of tuberculosis). They are
CC

CC	components.	
XX		
50	Sequence 433 AA;	
	Query Match	98.2%; Score 718; DB 20; Length 433;
	Best Local Similarity	98.6%; Pred. No. 1e-72;
	Matches 141; Conservative	0; Mismatches 2; Indels 0; Gaps 0
QY	1 ATTLPVQKHPRSLPEEFSLFAAFSPFAGLRTPDTRLMRLEDEKKEGRIYVRAELPGVD	60

OY	61	PDKQVHTAVRGGQTLTIAERFEKDLGRSEFAVGSFVRVYSLDPVGDEDIDKATYKGI	120
Db	70	pdkavdlmrvdqqltlikaertegkdfgrsefaygslyrvslpvgadeddlnaklykgl	129
OY	121	LTVSVAVSEGGKPETEKHQIIRSTN	143
Db	130	ltvsavsegkptekehqirstn	152
RESULT	5		
Y32064			
ID	Y32064	standard; Protein; 856 AA.	
XX			
AC	Y32064;		
XX			
DT	17-JAN-2000	(first entry)	
XX			
DE		Mycobacterium tuberculosis antigen fusion protein Mtb88f.	
XX			
KW		Tuberculosis; antigen; fusion protein; Mtb88f; ERD14; DPV; MTL;	
KW		MSL; mTCC2; diagnosis; therapy; vaccine; Immunogen.	
XX			
DS		Mycobacterium tuberculosis.	

```

XX Key Location/Qualifiers
FH Peptide 1..8
FT /note="Met/His tag"
FT 9..152
FT Protein /note="ERD14"
FT 155..236
FT Protein /note="DPV"
FT 239..332
FT Protein /note="MTI"
FT 335..431
FT Protein /note="MSL"
FT 434..856
FT Protein /note="MTTC2"
XX
XX MO951748-AZ.
XX
XX 14-OCT-1999.
XX
XX 07-APR-1999; 99MO-US07717.
XX
XX 07-APR-1998; 98US-0056556.
XX 30-DEC-1998; 98US-0223040.
XX (CORI-) CORIXA CORP.
XX
XX Skelky YAW, Alderson M, Campos-Neto A;
XX WPI: 1999-601610/51.
XX DR N-PSDB; 220199.
XX
XX New fusion proteins useful for diagnosis, prevention and treatment of
XX tuberculosis -
XX
XX Claim 1; Fig 6A-B; 83pp; English.
XX
XX This sequence represents a recombinant Mycobacterium tuberculosis
XX pent-antigen fusion protein, termed Mb88f, composed of the antigens
XX ERD14, DPV, MTI, MSL and MTTC2. The fusion protein is expressed in
XX host cells using a vector carrying a polynucleotide (see 220199)
XX comprising the coding sequences for the 5 antigens. The invention
XX provides fusion proteins (see Y32059-71) containing at least 2 M.
XX tuberculosis antigens. The new fusion proteins and polynucleotides
XX encoding them are useful as vaccines for preventing tuberculosis
XX (claimed), for diagnosis (via in vitro assays or intradermal skin
XX tests for detection of anti-M. tuberculosis antibodies), monitoring
XX of disease progression, and treatment of tuberculosis. They are
XX more effective immunogens than mixtures of the individual protein
XX components.
XX
XX SQ Sequence 856 AA:
XX
XX Query Match 98.2%; Score 718; DB 20; Length 856;
XX Best Local Similarity 98.6%; Pred. No. 2.8e-72;
XX Matches 141; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
XX 1 ATTLPVQRHRSILFPESESEFAAFSPAGLRPTEDTRLMLDEDMKRGREYRAELPGVD 60
XX |||||||
XX Db 10 attlpvqrhprslfpefsefafaafpsfaglrptctdlrmldeemkgrfeyrvaelpgvd 69
XX
XX 61 PDKDVHIMVRDGLTIKAERTEQKLDGRSEFAVGSFVRTVSLPVGADDDDIKATYDKGI 120
XX |||||||
XX Db 70 pdkdvdlmvrddglltkaerteqkldfgrsefaygsfvrvtvslpvgadeddikatydkgi 129
XX
XX 121 LTVSAVASEGKPTKHKIQRSTN 143
XX |||||||
XX Db 130 ltvsavasegkptekhkqrstn 152
XX
XX RESULT 6
XX R12489
XX ID R12489 standard; protein; 131 AA.

```

```

XX AC R12489;
XX
XX 10-SEP-1991 (first entry)
XX
XX Peptide derived from and antigenic to Mycobacterium tuberculosis.
XX
XX Tuberculosis; vaccine; lymphoproliferative; South Indian isolate S111.
XX
XX Mycobacterium tuberculosis.
XX
XX GB2239246-A.
XX
XX 26-JUN-1991.
XX
XX 17-DEC-1990; 90GB-0027267.
XX
XX 22-DEC-1989; 89IN-0000938.
XX 27-MAR-1990; 90SE-0001105.
XX
XX (ASTR ) ASTRA AB.
XX
XX Jagannath C, Balanesh M, Srinivasa BR;
XX WPI: 1991-187585/26.
XX
XX New protein and peptide(s) from Mycobacterium tuberculosis - for
XX PT early detection of tuberculosis and for treating tuberculosis
XX PT related inflammation by T-cell proliferation.
XX
XX Claim 1; Page 25; 35pp; English.
XX
XX The peptide and fragments thereof may be used in diagnosis, therapy
XX CC and prophylaxis of human and experimental tuberculosis. The product
XX CC is lymphoproliferative and may thus be used to stimulate human
XX CC peripheral blood lymphocytes for experimental use, and in the
XX CC production of cellular growth and differentiation factors, and
XX CC treatment of cancer by non-specifically boosting cellular immunity.
XX
XX SQ Sequence 131 AA:
XX
XX Query Match 48.7%; Score 356; DB 12; Length 131;
XX Best Local Similarity 55.3%; Pred. No. 1.1e-32;
XX Matches 89; Conservative 4; Mismatches 4; Indels 64; Gaps 7;
XX
XX 1 ATTLPVQRHRSILFPESESEFAAFSPAGLRPTEDTRLMLDEDMKRGREYRAELPGVD 52
XX |||||||
XX Db 1 attlpvqrhprslfpefsefafaafpsfaglrptctdlrmldeemkgrfeyrvaelpgvd 57
XX
XX 53 RAELPGVDPKDVHIMVRDGLTIKAERTEQKLDGRSEFAVGSFVRTVSLPVGADDDDI 112
XX |||
XX Db 58 ---lp-----vakhgcl-----rsefaygsfvrvtvslpvgadeddi 90
XX
XX 113 KATYDK-----GILTVSAV 127
XX :|||||
XX Db 91 ratcydkryevrddfgraelpgvddcvcitrglltvsavc 131
XX
XX RESULT 7
XX R85672
XX ID R85672 standard; peptide; 48 AA.
XX
XX R85672;
XX
XX 25-JAN-1996 (first entry)
XX
XX 12 KD M.tuberculosis extracellular protein N-terminal fragment.
XX
XX Majorly abundant extracellular product; MAP: M.tuberculosis;
XX immunise; vaccinating agent; vaccine; M.bovis; M.marinum; M.kansasii;
XX M.aviumintracellulare; M.fortuitum; M.chelonae; M.scrofulaceum;
XX M.leprae; M.africanum; M.ulcerans; M.microti; human; cat; dog;
XX

```


XX Harth G, Horwitz MA;
 PI
 XX
 DR WPI; 1997-020936/02.
 XX
 PT Vaccines derived from M.tuberculosis major abundant extracellular
 PT proteins - are easy to prepare and less toxic than conventional
 PT killed or attenuated vaccines, useful for protecting against or
 PT treating Mycobacterial infections
 XX
 PS Example 2; Page 107; 193pp; English.
 XX
 CC A vaccinating agent for promoting an immune response in a mammal
 CC against Mycobacterium pathogens comprises at least one
 CC immunodominant epitope of at least one majorly abundant
 CC extracellular protein, i.e. the M.tuberculosis 110, 80, 71, 58, 45,
 CC 32A, 32B, 30, 24, 23, 5, 23, 16, 14 or 12 kD proteins, or their
 CC analogues, homologues and subunits. The present sequence represents
 CC the N-terminal amino acid sequence of the 12 kD protein. The N-terminal
 CC amino acid sequences of the major abundant extracellular products were
 CC determined to provide structural data and to uncover possible
 CC relationships between the proteins. The vaccinating agents are used to
 CC protect against (or to treat existing) infections by Mycobacterium
 CC (especially M.tuberculosis) while the epitopes can also be used to
 CC detect presence of an immune response to a Mycobacterium pathogen. The
 CC vectors, containing the DNA for the extracellular proteins, are used to
 CC transform cells for production of recombinant DNA molecules. More
 CC generally the DNA from other pathogens can be used in vaccines, e.g.
 CC against other bacteria, viruses, fungi and protozoa. Since different
 CC combinations of DNA can be used, a wide range of effective compositions
 CC can be produced. They generate a response against the antigens most
 CC often found on infected cells during the infection, regardless of the
 CC strength or specificity of the immune response. The vaccines are easy to
 CC produce and less toxic than known killed or attenuated vaccines, so can
 CC be given to immunocompromised subjects, e.g. those with HIV infection.
 XX
 SQ Sequence 48 AA;
 XX
 Query Match 32.6%; Score 238; DB 18; Length 48;
 Best Local Similarity 97.9%; Pred. No. 5.3e-20;
 Matches 47; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 34 FDTLRMRLEDEMKREGRYEAELPGVDPKDVHIMVRDGLTKAERT 81
 ID 1 fdtlrmlrledemkgrgyevraelpgvdpkdvimvrdgltikaert 48
 DB
 XX
 ULT 10
 ID 570
 W75570 standard; protein; 48 AA.
 XX
 AC W75570;
 XX
 DT 23-OCT-1998 (first entry)
 XX
 DE M. tuberculosis 12 kD extracellular product N-terminal sequence.
 XX
 KM Mycobacterium tuberculosis; vaccination; extracellular product;
 KM immunodominant epitope; interleukin-12; MF59; immune response;
 KM opsonising humoral response; intracellular pathogen.
 XX
 OS Mycobacterium tuberculosis.
 OS
 PN WO9831388-A1.
 XX
 PD 23-JUL-1998.
 XX
 PF 15-JAN-1998; 98WO-US00942.
 XX
 PR 21-JAN-1997; 97US-0786533.
 XX
 PA (REGC) UNIV CALIFORNIA.

XX Harth G, Horwitz MA, Lee B;
 PI
 XX
 DR WPI; 1998-413815/35.
 XX
 PT Vaccines against Mycobacterium containing major extracellular
 PT proteins - used to, e.g. induce protective and therapeutic immune
 PT responses, and for detecting an immune response
 XX
 PS Example 2; Page 35; 236pp; English.
 XX
 CC Sequences shown in W75570 to W75586 represent N-terminal amino acid
 CC sequences of 14 exemplary major abundant extracellular products of
 CC Mycobacterium tuberculosis. The invention provides an agent for
 CC vaccinating mammals against Mycobacterium. The agent comprises at least
 CC one of the major abundant extracellular 110, 80, 71, 58, 45, 32A, 32B,
 CC 30, 24, 23, 5, 23, 16, 14 or 12 kDa proteins of M. tuberculosis, or at
 CC least 1 of their immunodominant epitopes and interleukin-12 (IL-12) or
 CC MF59 as adjuvants. The agent containing the nucleic acid encoding the
 CC extracellular products are used to raise a protective or therapeutic
 CC immune response against Mycobacterium, specifically M. tuberculosis.
 CC The immunodominant epitopes can also be used (typically in a cutaneous
 CC hypersensitivity test) to detect an immune response to vaccination.
 CC Preparation of the agent does not require selection of the most
 CC immunogenic products, so large scale production and purification are
 CC easy, resulting in a consistent, standardised formulation, having lower
 CC toxicity than killed or attenuated vaccines. The agents provide a rapid
 CC and effective response (including a strong cell-mediated component) and
 CC are safe even in immunocompromised subjects. They prevent development of
 CC an opsonising humoral response that might spread intracellular pathogens.
 XX
 SQ Sequence 48 AA;
 XX
 Query Match 32.6%; Score 238; DB 19; Length 48;
 Best Local Similarity 97.9%; Pred. No. 5.3e-20;
 Matches 47; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 34 FDTLRMRLEDEMKREGRYEAELPGVDPKDVHIMVRDGLTKAERT 81
 ID 1 fdtlrmlrledemkgrgyevraelpgvdpkdvimvrdgltikaert 48
 DB
 XX
 RESULT 11
 ID R43654
 R43654 standard; Protein; 181 AA.
 XX
 AC R43654;
 XX
 DT 16-MAY-1994 (first entry)
 XX
 DE c424 gene product.
 XX
 KM Production; high levels.
 XX
 OS Aspergillus nidulans FGSC NO. A89.
 OS
 PN JP05268964-A.
 XX
 PD 19-OCT-1993.
 XX
 PF 30-MAR-1992; 92JP-0074092.
 XX
 PR 30-MAR-1992; 92JP-0074092.
 XX
 PA (MEIP) MEIJI MILK PROD CO LTD.
 XX
 DR WPI; 1993-364289/46.
 DR N-PSDB; Q51347.
 XX
 PT Promoter of aspergillus used to produce animal protein - has
 PT sequence of 480 base pairs
 XX

PS Example: Page 5-6; 8pp; Japanese.

XX The sequence is that encoded by the c424 gene which was isolated
CC from an *Aspergillus nidulans* cDNA library.

XX Sequence 181 AA;

Query Match 15.5%; Score 113.5; DB 14; Length 181;
Best Local Similarity 25.8%; Pred. No. 3.8e-05;
Matches 47; Conservative 20; Mismatches 46; Indels 69; Gaps 9;

OY 16 EFSELP-----AAPSAGLRPFDFRLMRLEDEMEKGRVAVAEIPGVD 60
DB 12 diaplfrllldydnhsarghsvgsfa--pfdr-----esnea-yhldqelpg1- 60
OY 61 PDKDVHINWRDQ-LTIKA-----ERTEQKDLDR----- 89
DB 61 pgsndietftqqlvkgfserereyhssddknqdqatengargessevaktgexvst 120
90 -----SEFAGSVRTVSLFVGADDEDDIKATYKGIITVSVAVSEKPTKHIQ 138
DB 121 kkaankxywsversvgeftgftfptrvngdvdvksldgllsl-vvpkavptakklt 179
OY 139 IR 140
DB 180 iq 181

RESULT 12

ID P50065 standard; protein; 153 AA.

XX P50065;

XX 12-NOV-1991 (first entry)

DE Soybean heat shock protein hs6871.

XX Soybean; heat shock protein gene hs6871; plasmid; herbicide;

XX Glycine max.

XX EPI59884-A.

XX 30-OCT-1985.

PD 12-APR-1985; 85EP-0302593.

XX 13-APR-1984; 84US-0599993.

XX (AGRI-) AGRIGENETICS RESEARCH ASSOCIATES LTD.

XX Key JL, Gurley WB, Nagao RT, Schoeffl F, Czarnicka E;

XX WPI: 1985-271444/44.

XX N-PSDB; NS0077.

PT Recombinant plasmid contg. plant DNA - controlling expression of heat

XX shock gene, inserted into tumour inducing plasmid of *agrobacterium*.

XX Disclosure; Fig. 4; 47pp; English.

CC DNA encoding soybean heat shock protein gene hs6871 is inserted

CC into a T-DNA fragment of A Ti plasmid from *Agrobacterium* and used in a

CC recombinant plasmid which also comprises a vector and opt. a

CC structural gene under the control of the soybean DNA. The vector is

CC esp. pBR322 and the structural gene is, eg the z gene encoding beta

CC galactosidase, or an endotoxin gene, or it is a soybean gene, eg for

CC herbicide resistance. The new plasmid can be used to control prodn.

CC (from the structural gene) of crystalline endotoxin (esp. from

CC *Bacillus thuringiensis*) or to introduce resistance to herbicides or

CC pathogens. The structural gene can also be used to recognise plant

CC cells which have been transformed with T-DNA.

XX Sequence 153 AA;

Query Match 15.2%; Score 111; DB 6; Length 153;
Best Local Similarity 27.5%; Pred. No. 5.6e-05;
Matches 44; Conservative 23; Mismatches 51; Indels 42; Gaps 7;

OY 12 SLPEPSELFAPPSFAGLRPT-FDFRLMRLEDEME----- 47
DB 2 slpsf-----fgrrsvfdpfsldvwdpfdftvpslsensafstlvdwke 52
OY 48 --GRTVRAELPGVDKDVHINWRDQ-LTIKAERTEOK---DLGRSEFAYGSFVRP 100
DB 53 tpeahvfkadipglk-ekkleigqgrylqsgservedndtwhvrsrklvr 111
OY 101 VSLPVGADDEDDIKATYKGIITVSVAVSE-GRPTKHIQ 139
DB 112 flipenakvdeykasmenvgllvlpkkelkprdvakid 151

RESULT 13

ID P50066 standard; protein; 154 AA.

XX P50066;

XX 12-NOV-1991 (first entry)

DE Soybean heat shock protein pe2019.

XX Soybean; heat shock protein gene pe2019; plasmid; herbicide;

XX Glycine max.

XX EPI59884-A.

XX 30-OCT-1985.

PD 12-APR-1985; 85EP-0302593.

XX 13-APR-1984; 84US-0599993.

XX (AGRI-) AGRIGENETICS RESEARCH ASSOCIATES LTD.

XX Key JL, Gurley WB, Nagao RT, Schoeffl F, Czarnicka E;

XX WPI: 1985-271444/44.

XX N-PSDB; NS0070.

PT Recombinant plasmid contg. plant DNA - controlling expression of heat

XX shock gene, inserted into tumour inducing plasmid of *agrobacterium*.

XX Disclosure; Fig. 1; 47pp; English.

CC DNA encoding soybean heat shock protein gene pe2019 is inserted

CC into a T-DNA fragment of A Ti plasmid from *Agrobacterium* and used in a

CC recombinant plasmid which also comprises a vector and opt. a

CC structural gene under the control of the soybean DNA. The vector is

CC esp. pBR322 and the structural gene is, eg the z gene encoding beta

CC galactosidase, or an endotoxin gene, or it is a soybean gene, eg for

CC herbicide resistance. The new plasmid can be used to control prodn.

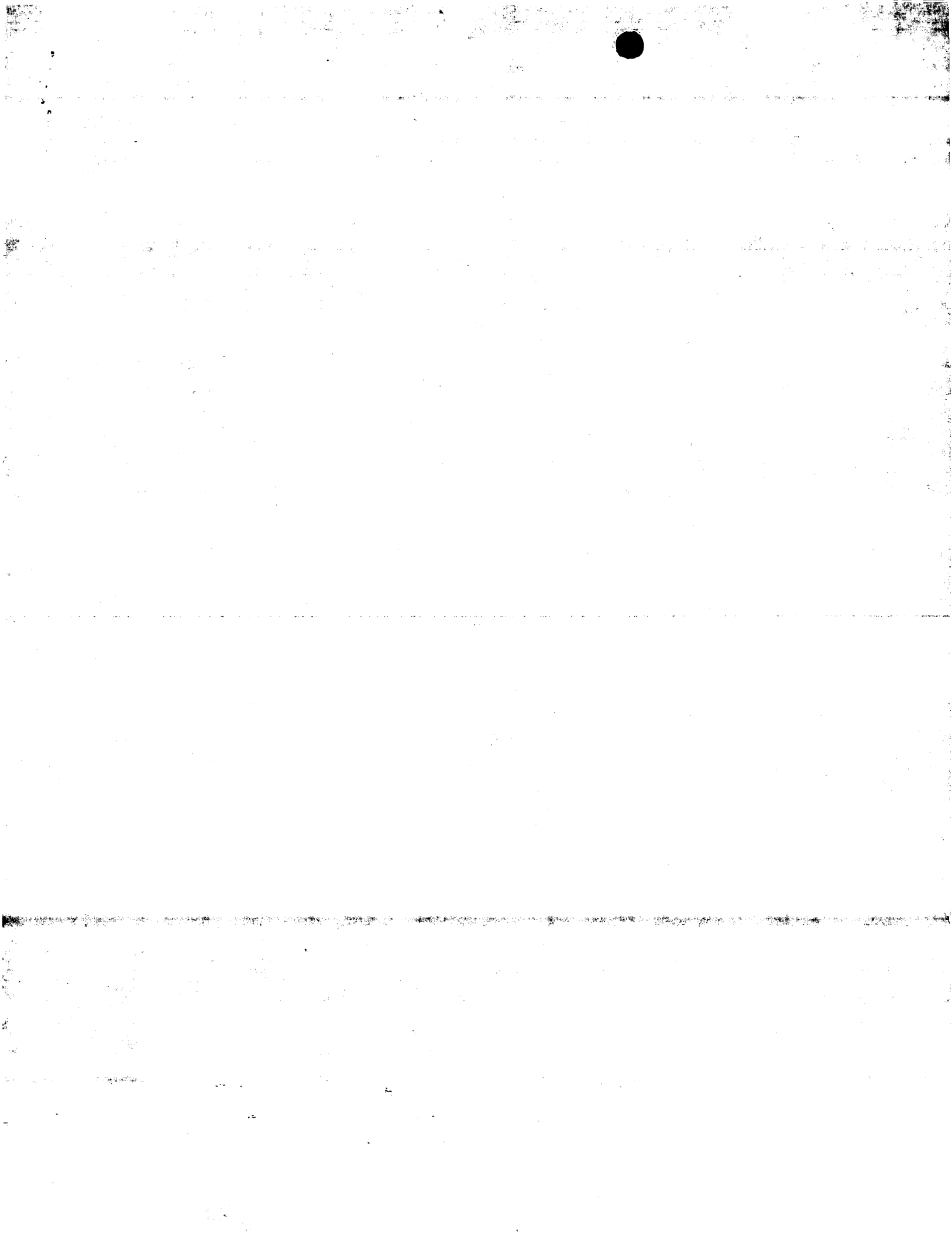
CC (from the structural gene) of crystalline endotoxin (esp. from

CC *Bacillus thuringiensis*) or to introduce resistance to herbicides or

CC pathogens. The structural gene can also be used to recognise plant

CC cells which have been transformed with T-DNA.

Query Match 14.3%; Score 104.5; DB 6; Length 154;
Best Local Similarity 33.3%; Pred. No. 0.00031;



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 12, 2001, 18:01:41 ; Search time 3358.72 Seconds
(without alignments)
2422.260 Million cell updates/sec

Title: US-09-461-774-9
Perfect score: 1161
Sequence: 1 aatcgcgcgcagtcgccgc.....cgccgcgagtaagtttga 1161

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 7991742 segs, 3503743858 residues
Total number of hits satisfying chosen parameters: 15983484

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: em_est40:*
42: em_estfun:*
43: em_esthum1:*

44: em_esthum2:*
45: em_esthum3:*
46: em_esthum4:*
47: em_esthum5:*
48: em_esthum6:*
49: em_esthum7:*
50: em_esthum8:*
51: em_esthum9:*
52: em_esthum10:*
53: em_esthum11:*
54: em_esthum12:*
55: em_esthum13:*
56: em_esthum14:*
57: em_esthum15:*
58: em_esthum16:*
59: em_esthum17:*
60: em_esthum18:*
61: em_esthum19:*
62: em_esthum20:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estov1:*
68: em_estov2:*
69: em_estp11:*
70: em_estp12:*
71: em_estp13:*
72: em_estp14:*
73: em_estp15:*
74: em_estro1:*
75: em_estro2:*
76: em_estro3:*
77: em_estro4:*
78: em_estro5:*
79: em_estro6:*
80: em_estro7:*
81: em_estro8:*
82: em_estro9:*
83: em_estro10:*
84: em_estro11:*
85: em_estro12:*
86: em_estro13:*
87: gb_est41:*
88: gb_est42:*
89: gb_est43:*
90: gb_est44:*
91: gb_est45:*
92: gb_est46:*
93: gb_est47:*
94: gb_est48:*
95: gb_est49:*
96: gb_est50:*
97: gb_est51:*
98: gb_est52:*
99: gb_est53:*
100: gb_est54:*
101: gb_est55:*
102: gb_est56:*
103: gb_est57:*
104: gb_est67:*
105: gb_est68:*
106: gb_est69:*
107: gb_est70:*
108: gb_est71:*
109: gb_est72:*
110: gb_est73:*
111: gb_est74:*
112: em_esthum21:*
113: em_esthum22:*
114: em_esthum23:*
115: em_estom1:*
116: em_estom2:*

117: em_estp16:*
 118: em_estp17:*
 119: em_estp18:*
 120: em_estro14:*
 121: em_estro15:*
 122: em_estro16:*
 123: em_estro17:*
 124: em_estro18:*
 125: em_estro19:*
 126: gb_est58:*
 127: gb_est59:*
 128: gb_est60:*
 129: gb_est61:*
 130: gb_est62:*
 131: gb_est63:*
 132: gb_est64:*
 133: gb_est65:*
 134: gb_est66:*
 135: gb_est67:*
 136: gb_est68:*
 137: gb_est69:*
 138: gb_est70:*
 139: gb_est71:*
 140: gb_est72:*
 141: gb_est73:*
 142: gb_est74:*
 143: gb_est75:*
 144: gb_est76:*
 145: gb_est77:*
 146: gb_est78:*
 147: gb_est79:*
 148: gb_est80:*
 149: gb_est81:*
 150: gb_est82:*
 151: gb_est83:*
 152: gb_est84:*
 153: gb_est85:*
 154: gb_est86:*
 155: gb_est87:*
 156: gb_est88:*
 157: gb_est89:*
 158: gb_est90:*
 159: gb_est91:*
 160: gb_est92:*
 161: gb_est93:*
 162: gb_est94:*
 163: gb_est95:*
 164: gb_est96:*
 165: gb_est97:*
 166: gb_est98:*
 167: gb_est99:*
 168: gb_est100:*
 169: gb_est101:*
 170: gb_est102:*
 171: gb_est103:*
 172: gb_est104:*
 173: gb_est105:*
 174: gb_est106:*
 175: gb_est107:*
 176: gb_est108:*
 177: gb_est109:*
 178: gb_est110:*
 179: gb_est111:*
 180: gb_est112:*
 181: gb_est113:*
 182: gb_est114:*
 183: gb_est115:*
 184: gb_est116:*
 185: gb_est117:*
 186: gb_est118:*
 187: gb_est119:*
 188: gb_est120:*
 189: gb_est121:*

190: gb_gss25:*
 191: gb_gss26:*
 192: gb_gss27:*
 193: gb_gss28:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	183.8	15.8	1382	94	AW731194	AW731194 GA_Ea001
2	174.6	15.0	1016	192	CNS03L09	AL249930 Tetradon
3	167.8	14.5	1246	94	AW727499	AW727499 GA_Ea001
4	155.8	13.4	1099	94	AW731227	AW731227 GA_Ea001
5	152.2	13.1	1097	94	AW731157	AW731157 GA_Ea001
6	152	13.1	1347	94	AW731204	AW731204 GA_Ea001
7	149	12.8	1711	94	AW731135	AW731135 GA_Ea001
8	148.2	12.8	1338	94	AW731160	AW731160 GA_Ea001
9	148	12.7	1509	108	BE455283	BE455283 HVSMEH009
10	147.8	12.7	1315	94	AW731134	AW731134 GA_Ea001
11	146.6	12.6	827	176	A2183795	A2183795 SP_1002.A
12	146.4	12.6	1292	94	AW727529	AW727529 GA_Ea001
13	146.2	12.6	1307	94	AW731210	AW731210 GA_Ea001
14	146	12.6	1440	94	AW727557	AW727557 GA_Ea001
15	145.8	12.6	1101	190	CNS0153F	AL104949 Drosophila
16	145.8	12.6	1618	94	AW727506	AW727506 GA_Ea001
17	145.4	12.5	1411	94	AW731189	AW731189 GA_Ea001
18	145.2	12.5	1303	94	AW727501	AW727501 GA_Ea001
19	144.4	12.4	1060	94	AW727527	AW727527 GA_Ea001
20	144	12.4	952	94	AW727551	AW727551 GA_Ea001
21	143.8	12.4	1468	94	AW731223	AW731223 GA_Ea001
22	143	12.3	1363	94	AW731198	AW731198 GA_Ea001
23	140.6	12.1	1006	134	BE034822	BE034822 ML04H02 M
24	140.4	12.1	1126	94	AW731180	AW731180 GA_Ea001
25	140.2	12.1	1086	94	AW727532	AW727532 GA_Ea001
26	139.6	12.0	821	172	A0893474	A0893474 HS_4832.A
27	139	12.0	1049	94	AW727475	AW727475 GA_Ea001
28	138.4	11.9	1397	94	AW727486	AW727486 GA_Ea001
29	138	11.9	1180	94	AW727502	AW727502 GA_Ea001
30	137.6	11.9	1674	94	AW731181	AW731181 GA_Ea001
31	137.4	11.8	1262	94	AW727556	AW727556 GA_Ea001
32	137.4	11.8	1328	108	BE455227	BE455227 HVSMEH009
33	137.2	11.8	1501	94	AW731188	AW731188 GA_Ea001
34	137.2	11.8	1622	94	AW731133	AW731133 GA_Ea001
35	136.6	11.8	1252	94	AW727555	AW727555 GA_Ea001
36	136.4	11.7	1141	94	AW731203	AW731203 GA_Ea001
37	136.4	11.7	1519	94	AW731138	AW731138 GA_Ea001
38	136.2	11.7	1188	108	BE455091	BE455091 HVSMEH009
39	135.4	11.7	918	172	A0895329	A0895329 HS_4832.A
40	134.8	11.6	1158	94	AW727550	AW727550 GA_Ea001
41	134	11.5	915	176	A2195437	A2195437 SP_1030.A
42	134	11.5	1210	94	AW731144	AW731144 GA_Ea001
43	133.8	11.5	1080	108	BE455154	BE455154 HVSMEH009
44	133.8	11.5	1081	94	AW727480	AW727480 GA_Ea001
45	133.2	11.5	1063	94	AW727530	AW727530 GA_Ea001

ALIGNMENTS

RESULT 1
 AW731194
 LOCUS
 DEFINITION GA_Ea0010K03 Gossypium arboreum 7-10 dpa fiber library Gossypium
 arboreum cDNA clone GA_Ea0010K03, mRNA sequence.
 ACCESSION AW731194
 VERSION AW731194.1 GI:7628853
 KEYWORDS EST.
 SOURCE Gossypium arboreum.

Db	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466	467	468	469	470	471	472	473	474	475	476	477	478	479	480	481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500	501	502	503	504	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520	521	522	523	524	525	526	527	528	529	530	531	532	533	534	535	536	537	538	539	540	541	542	543	544	545	546	547	548	549	550	551	552	553	554	555	556	557	558	559	560	561	562	563	564	565	566	567	568
----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

JOURNAL COMMENT

Unpublished (2000)
Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu

High quality sequence stop: 1099.
Location/Qualifiers
1. 1099
/organism="Gossypium arboreum"
/strain="AKA"
/cultivar="8400"
/db_xref="taxon:29729"
/clone="GA_Ea0010021"
/clone_idb="Gossypium arboreum 7-10 dpa fiber library"
/tissue_type="Fibers isolated from bolls harvested 7-10 dpa"
/lab_host="E. coli"
/note="vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"

BASE COUNT

50 a 626 c 363 g 57 t 3 others

ORIGIN

[illegible]

Db 828 TACCCCTCCGCTCCGCCCCCCCCCCCCCCCCCGCCACCCCGCCGCGCCGCCGCCCA 769
Oy 647 cgaacacccgctgagacgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 706
Db 768 CCCCCGCTCCCG 709
Oy 707 ccccgacgtatgcacacgagctcgaacacgagcgagcgagcgagcgagcgagcgagcg 766
Db 708 CCGGACCGCGGACACCG 650
Oy 767 ctgagcgagtgctccgagctcgcgagcgagcgagcgagcgagcgagcgagcgagcg 826
Db 649 CCG 590
Oy 827 cgc 886
Db 589 CACAGCCCG 530
887 cgc 946
529 GAGCGCGCGCGGAAACG-CGACGCGCGCGGACACAGCTCCGCGCGCTCGCGCGCGCG 471
Oy 947 ccccgagacgc 1006
Db 470 CCCCACCTCCCG 411
Oy 1007 cgc 1031
Db 410 CGCTCCGCGCGCGCGCGCTCTGCGCTCC 386

RESULT 10
AW731134
LOCUS AW731134 1315 bp mRNA
DEFINITION GA_Ea0010A03 Gossypium arboreum 7-10 dpa fiber library Gossypium
arborescens cDNA clone GA_Ea0010A03, mRNA sequence.
VERSION AW731134
KEYWORDS AW731134.1 GI:7628791
SOURCE EST.
ORGANISM Gossypium arboreum.
REFERENCE
AUTHORS Leslie, A., Frisch, D., Yu, Y., Wood, T. C., Wang, R. A. and Wilkins, T. A.
TITLE An integrated analysis of the genetics, development, and evolution
of the cotton fiber
JOURNAL Unpublished (2000)
COMMENT Contact: Ming RA
Clemson University Genomics Institute
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
High quality sequence stop: 1315.
FEATURES
source
1. .1315
/organism="Gossypium arboreum"
/strain="AKA"
/cultivar="8400"
/db_xref="taxon:29729"
/clone="GA_Ea0010A03"
/tissue_type="Fibers isolated from bolls harvested 7-10
dpa"
/lab_host="E. coli"
/note="Vector: PBK-CMV; Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 78 a 777 c 348 g 101 t 11 others
ORIGIN

Query Match 12.7%; Score 147.8; DB 94; Length 1315;
Best Local Similarity 49.7%; Pred. No. 1.6e-15;
Matches 483; Conservative 0; Mismatches 483; Indels 6; Gaps 4;

Oy 21 ggcgcacacagc 80
Db 329 GGGCG 388
Oy 81 ggc 140
Db 389 CCCCCG 447
Oy 141 ggc 200
Db 448 GCG 507
Oy 201 ggc 260
Db 508 CCG 567
Oy 261 ggc 320
Db 568 GCG 626
Oy 321 tcgc 380
Db 627 CCG 686
Oy 381 ggc 440
Db 687 GCG 746
Oy 441 ggc 500
Db 747 CCG 806
Oy 501 ggc 560
Db 807 CCG 866
Oy 561 ggc 620
Db 867 GCG 924
Oy 621 ggc 680
Db 925 TCCG 984
Oy 681 ggc 740
Db 985 GCG 1044
Oy 741 ggc 800
Db 1045 GCG 1104
Oy 801 ggc 860
Db 1105 GCG 1164
Oy 861 ggc 920
Db 1165 GCG 1222
Oy 921 acctgc 980
Db 1223 GCG 1282
Oy 981 ggc 992
Db 1283 TCG 1294

```

RESULT 11
LOCUS A2183795/c 827 bp DNA GSS 30-AUG-2000
DEFINITION SP_1002.A1.A06_Sp6 Strongylocentrotus purpuratus, purple sea urchin
            , sperm genomic BAC library Strongylocentrotus purpuratus genomic
            clone plate=1002 Col-11 Row=A, DNA sequence.
ACCESSION A2183795
VERSION A2183795.1 GI:8356170
KEYWORDS GSS.
SOURCE Strongylocentrotus purpuratus.
          Strongylocentrotus purpuratus
          Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
          Echinoidea; Echinozoa; Echinodermata; Echinodermata; Echinodermata;
          Strongylocentrotidae; Strongylocentrotus.
REFERENCE 1 (bases 1 to 827)
          Cameron,R.A., Mahliras,G., Rast,J.P., Martinez,P., Biondi,T.R.,
          Swartzell,S., Wallace,J.C., Poustka,A.J., Livingston,B.T., Wray
          ,G.A., Eitensohn,C.A., Lehrach,H., Britten,R.J., Davidson,E.H. and
          Hood,L.
          A sea urchin genome project: Sequence scan, virtual map, and
          additional resources
          Proc. Natl. Acad. Sci. U. S. A. 97 (17), 9514-9518 (2000)
JOURNAL 20402566
MEDLINE Contact: Cameron, RA, Davidson, EH, Hood, L
          Division of Biology 156-29
          California Institute of Technology
          Pasadena California 91125, USA
          Tel: (626) 395-8421
          Fax: (626) 793-3047
          Email: acameron@caltech.edu
          Plate: 1002 row: A column: 11
          Seq primer: SP6
          Class: BAC ends
          High quality sequence stop: 827.
FEATURES
    source
        1..827
        location/Qualifiers
            /organism="Strongylocentrotus purpuratus"
            /db_xref="taxon:7668"
            /clone="Plate=1002 Col=11 Row=A"
            /clone_lib="Strongylocentrotus purpuratus, purple sea
            urchin, sperm genomic BAC library"
            /note="Organ: sperm; Vector: BAC3.6; BAC Clones in E-Coli
            DH10B"
BASE COUNT 20 a 166 c 604 g 13 t 24 others
ORIGIN
Query Match 12.6%; Score 146.6; DB 176; Length 827;
Best Local Similarity 51.0%; Pred. No. 2,7e-15;
Matches 369; Conservative 0; Mismatches 350; Indels 5; Gaps 2;
QY 295 ccggagtgcgagcgcgagaccggttcgagtgcgagtgccgcaatgccgacgagccgaag 354
    || || || || || || || || || || || || || || || || || || || || ||
DB 808 ccccccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 749
QY 355 aggatgacgagtgctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 414
    || || || || || || || || || || || || || || || || || || || || ||
DB 748 ccnncnccgnggnggnggnggnggnggnggnggnggnggnggnggnggnggnggnggngg 689
QY 415 gcggtgacagtgctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 474
    || || || || || || || || || || || || || || || || || || || || ||
DB 688 cccgcgncnccgcccccccgccgccccgccccgccccgccccgccccgccccgccccgcccc 629
QY 475 cctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 530
    || || || || || || || || || || || || || || || || || || || || ||
DB 628 ccccgccnccgcccccccgccgccccgccccgccccgccccgccccgccccgccccgcccc 569
QY 531 gcccgagacgcacattggagagacagccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 590
    || || || || || || || || || || || || || || || || || || || || ||
DB 568 gccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 509
QY 591 gcccccgctgacgtcaaccgcgcgcgtgacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 650

```

```

DB 508 gccccccgcnccgcccccccccccccccccccccccccccccccccccccccccccccccc 449
QY 651 cagcccgctgacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 710
    || || || || || || || || || || || || || || || || || || || || ||
DB 448 ccccgccnccgcccccccgccgccccgccccgccccgccccgccccgccccgccccgccccgcccc 389
QY 711 gccgatgcacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 770
    || || || || || || || || || || || || || || || || || || || || ||
DB 388 cccccccgcnccgcccccccccccccccccccccccccccccccccccccccccccccccc 330
QY 771 gccggtgttcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 830
    || || || || || || || || || || || || || || || || || || || || ||
DB 329 ccccgccgcccccccgccgccccgccccgccccgccccgccccgccccgccccgccccgcccc 270
QY 831 ggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 890
    || || || || || || || || || || || || || || || || || || || || ||
DB 269 cccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 210
QY 891 cccgcgcgttgacgatagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 950
    || || || || || || || || || || || || || || || || || || || || ||
DB 209 cgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 150
QY 951 ggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1010
    || || || || || || || || || || || || || || || || || || || || ||
DB 149 gacgctcccccccccccccccccccccccccccccccccccccccccccccccccccccccc 90
QY 1011 cccc 1014
    || || || || || || || || || || || || || || || || || || || || ||
DB 89 cccc 86

```

```

RESULT 12
LOCUS AW727529 1292 bp mRNA EST 20-APR-2000
DEFINITION GA_Ea0012M14 Gossypium arboreum 7-10 dpa fiber library Gossypium
            arboreum cDNA clone GA_Ea0012M14, mRNA sequence.
ACCESSION AW727529
VERSION AW727529.1 GI:7625085
KEYWORDS EST.
SOURCE Gossypium arboreum.
          Gossypium arboreum
          Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
          Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
          Malvales; Malvaceae; Gossypium.
          1 (bases 1 to 1292)
          Leslie,A., Frisch,D., Yu,Y., Wood,T.C., Wing,R.A. and Wilkins,T.A.
          An integrated analysis of the genetics, development, and evolution
          of the cotton fiber
          Unpublished (2000)
REFERENCE 1
          Clemson University Genomics Institute
          Clemson University
          100 Jordan Hall, Clemson, SC 29634, USA
          Tel: 864 656 7288
          Fax: 864 656 4293
          Email: rwing@clemson.edu
          High quality sequence stop: 1292.
FEATURES
    source
        1..1292
        location/Qualifiers
            /organism="Gossypium arboreum"
            /strain="AKA"
            /cultivar="8400"
            /db_xref="taxon:29729"
            /clone="GA_Ea0012M14"
            /clone_lib="Gossypium arboreum 7-10 dpa fiber library"
            /tissue_type="Fibers isolated from bolls harvested 7-10
            dpa"
            /lab_host="E. coli"
            /note="Vector: pBR-CMW; Site_1: EcoRI; Site_2: XhoI"
BASE COUNT 59 a 761 c 370 g 89 t 13 others
ORIGIN

```


[illegible]

RESULT	14
LOCUS	AW727557
DEFINITION	AW727557 1440 bp mRNA EST 20-APR-2000
ACCESSION	GA_Ea0012022 Gossypium arboreum 7-10 dpa fiber library Gossypium
VERSION	arboreum cDNA clone GA_Ea0012022, mRNA sequence.
KEYWORDS	AW727557.1 GI:7625114
SOURCE	EST.
ORGANISM	Gossypium arboreum.
TITLE	Gossypium arboreum
REFERENCE	Eukaryota: Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta
AUTHORS	Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II
COMMENT	Malvales; Malvaceae; Gossypium.
JOURNAL	1 (bases 1 to 1440)
CONTACT	Leslie, A., Fritsch, D., Yu, Y., Wood, T. C., Wing, R. A. and Wilkins, T. A.
UNPUBLISHED	An integrated analysis of the genetics, development, and evolution
CONTACT	of the cotton fiber
CONTACT	Unpublished (2000)
CONTACT	Contact: Wing RA
CONTACT	Clemson University Genomics Institute
CONTACT	Clemson University
CONTACT	100 Jordan Hall, Clemson, SC 29634, USA
CONTACT	Tel: 864 656 7288
CONTACT	Fax: 864 656 4293
CONTACT	Email: rwing@clemson.edu
FEATURES	High quality sequence stop: 1440.
SOURCE	Location/Qualifiers
SOURCE	1..1440

```

/organism="Gossypium arboreum"
/strain="AKA"
/cultivar="8400"
/db_xref="taxon:29729"
/clone="Ga_Ea0012022"
/clone_lib="Gossypium arboreum 7-10 dpa fiber library"
/tissue_type="Fibers isolated from bolls harvested 7-10 dpa"
/lab_host="E. coli"
note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"

```

BASE COUNT	83 a	701 c	556 g	75 t	25 others
ORIGIN					
Query Match		12.6%	Score 146;	DB 94;	Length 1440;
Best Local Similarity		51.9%	Pred. No. 3.2e-15;		
Matches 522; Conservative		0;	Mismatches 467;	Indels 17;	Gaps 8;

[illegible]

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 12, 2001, 20:14:22 ; Search time 153.29 Seconds

(Without alignments)
1220.609 Million cell updates/sec

Title: US-09-461-774-9

Sequence: 1 aatccagacccagtcgcgc.....cgcgcagtaagtttga 1161

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 280836 seqs, 80580151 residues

Optimal number of hits satisfying chosen parameters: 561672

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: /cgn2_6/prodata/1/ina/5A_COMB.seq.*
2: /cgn2_6/prodata/1/ina/5B_COMB.seq.*
3: /cgn2_6/prodata/1/ina/6_COMB.seq.*
4: /cgn2_6/prodata/1/ina/PCRTUS_COMB.seq.*
5: /cgn2_6/prodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	189.4	16.3	6530	2 US-08-146-930-1	Sequence 1, Appl 1
C 2	189.4	16.3	6530	2 US-08-458-240-1	Sequence 1, Appl 1
C 3	189.4	16.3	6530	4 PCT-US93-03993-1	Sequence 1, Appl 1
C 4	128.4	11.1	2064	1 US-08-343-428-1	Sequence 1, Appl 1
C 5	126	10.9	4257	2 US-08-690-473-1	Sequence 1, Appl 1
C 6	126	10.9	12001	1 US-08-458-568A-11	Sequence 11, Appl 1
C 7	121.8	10.5	38506	3 US-09-320-878-19	Sequence 19, Appl 1
C 8	121.6	10.5	15239	1 US-08-390-878-17	Sequence 17, Appl 1
C 9	114.4	9.9	8438	1 US-07-945-283-1	Sequence 1, Appl 1
C 10	113.2	9.8	44377	2 US-08-804-227C-7	Sequence 7, Appl 1
C 11	113.2	9.8	44377	2 US-08-804-198-1	Sequence 1, Appl 1
C 12	112.2	9.7	774	3 US-08-956-307B-12	Sequence 12, Appl 1
C 13	112.2	9.7	778	3 US-08-956-307B-11	Sequence 11, Appl 1
C 14	109.8	9.5	44377	2 US-08-804-227C-7	Sequence 7, Appl 1
C 15	109.8	9.5	44377	2 US-08-804-198-1	Sequence 1, Appl 1
C 16	108.4	9.3	43280	2 US-08-804-227C-1	Sequence 1, Appl 1
C 17	107.6	9.3	2721	5 5215881-2	Patent No. 5215881
C 18	107.2	9.2	6192	2 US-08-479-537A-1	Sequence 1, Appl 1
C 19	107.2	9.2	6449	2 US-08-479-537A-4	Sequence 4, Appl 1
C 20	106	9.1	8438	1 US-07-945-283-1	Sequence 1, Appl 1
C 21	104.4	9.0	12001	1 US-08-458-568A-11	Sequence 11, Appl 1
C 22	104	8.8	2064	1 US-08-343-428-1	Sequence 1, Appl 1
C 23	102.6	8.8	756	1 US-08-642-255-50	Sequence 50, Appl 1
C 24	101.4	8.7	5382	2 US-08-403-852D-1	Sequence 1, Appl 1
C 25	101.4	8.7	5382	2 US-08-510-646B-1	Sequence 1, Appl 1
C 26	101.2	8.7	4257	3 US-08-690-473-1	Sequence 1, Appl 1
C 27	99	8.5	1931	2 US-09-130-114-2	Sequence 2, Appl 1
C 28	98.2	8.5	28958	1 US-08-258-261B-6	Sequence 6, Appl 1

29	98.2	8.5	28958	1 US-08-456-837-6	Sequence 6, Appl 1
30	98.2	8.5	28958	1 US-08-457-342-6	Sequence 6, Appl 1
31	98.2	8.5	28958	1 US-08-457-646A-6	Sequence 6, Appl 1
32	98.2	8.5	28958	1 US-08-458-076A-6	Sequence 6, Appl 1
33	98.2	8.5	28958	1 US-08-764-233A-4	Sequence 4, Appl 1
34	98.2	8.5	28958	1 US-08-457-335A-6	Sequence 6, Appl 1
35	98.2	8.5	28958	1 US-08-729-214-6	Sequence 6, Appl 1
36	98.2	8.5	28958	3 US-09-028-934-6	Sequence 6, Appl 1
37	98.2	8.5	49377	1 US-08-764-233A-1	Sequence 1, Appl 1
C 38	98	8.4	1288	1 US-08-440-856A-9	Sequence 9, Appl 1
C 39	97.6	8.4	833	2 US-08-403-852D-3	Sequence 3, Appl 1
C 40	97.6	8.4	833	3 US-08-510-646B-3	Sequence 3, Appl 1
C 41	97.6	8.4	18994	1 US-08-459-586-4	Sequence 4, Appl 1
C 42	97.6	8.4	18994	2 US-08-282-696-4	Sequence 4, Appl 1
C 43	96.6	8.3	11219	1 US-07-642-734C-1	Sequence 1, Appl 1
C 44	96.6	8.3	11219	3 US-08-439-009A-1	Sequence 1, Appl 1
C 45	96.2	8.3	12588	2 US-08-387-942C-1	Sequence 1, Appl 1

ALIGNMENTS

RESULT 1
US-08-146-930-1/c
Sequence 1, Application US/08146930
Patent No. 5958764
GENERAL INFORMATION:
APPLICANT: Roop, Dennis R.
APPLICANT: Rothenberg, David A.
TITLE OF INVENTION: SPECIFIC EXPRESSION VECTORS
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: LYON & LYON
STREET: 611 West Sixth Street
City: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90017
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 MB storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: IBM MS-DOS (Version 5.0)
SOFTWARE: WordPerfect (Version 5.1)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/146,930
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA: including application
PRIOR APPLICATION DATA: described below: two
APPLICATION NUMBER: 07/876,286
FILING DATE: April 30, 1992
APPLICATION NUMBER: No. 5958764 yet assigned (204/132)
ATTORNEY/AGENT INFORMATION:
NAME: Warburg, Richard J.
REGISTRATION NUMBER: 32,327
REFERENCE/DOCKET NUMBER: 204/152
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 6530 bases
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-146-930-1

[illegible]

Oy	1045	gagcgcgcccgagaacagctgggtl99gcytlttcacaatlglttagagcaggccctccaaagcgc	1104
Db	28550	GCACGACCCTGGACAACAGCAGCGCGTCCGCCGCCGCCGCCGTCCCTCGTGAGGAGCACCAATGCG	28491
Oy	1105	gccgcgagcgagcgagcctcagcgtctgcgtlacgcgcgcagcgccccgcagtaaaag	1155
Db	28490	CCGTCGGGGCCGGGAACGCGCGGCAAGTGGGCGTGGCGCGCTCGGCCAAG	28440
RESULT 15			
US-08-804-198-1/c			
Sequence 1, Application US/08804198			
Patent No. 5945320			
GENERAL INFORMATION:			
APPLICANT: Burgett, Stanley G.			
APPLICANT: Kuhstoss, Stuart A.			
APPLICANT: Rao, Nagaraja R.			
APPLICANT: Richardson, Mark A.			
APPLICANT: Rosteck, Paul R., Jr.			
TITLE OF INVENTION: PLATENOLIDE SYNTHASE GENE			
NUMBER OF SEQUENCES: 6			
CORRESPONDENCE ADDRESSES:			
ADDRESSEE: PAUL R. CANTRELL 1138			
STREET: LILLY CORPORATE CENTER			
CITY: INDIANAPOLIS			
STATE: IN			
COUNTRY: USA			
ZIP: 46285			
COMPUTER READABLE FORM:			
MEDIUM TYPE: Floppy disk			
OPERATING SYSTEM: Macintosh 7.0			
SOFTWARE: Microsoft Word 5.1			
CURRENT APPLICATION DATA:			
APPLICATION NUMBER: US/08/804,198			
FILING DATE:			
CLASSIFICATION: 435			
ATTORNEY/AGENT INFORMATION:			
NAME: CANTRELL, PAUL R.			
REGISTRATION NUMBER: 36,470			
REFERENCE/DOCKET NUMBER: P113			
TELECOMMUNICATION INFORMATION:			
TELEPHONE: 317-276-3885			
INFORMATION FOR SEQ ID NO: 1:			
SEQUENCE CHARACTERISTICS:			
LENGTH: 44377 base pairs			
TYPE: nucleic acid			
STRANDEDNESS: single			
TOPOLOGY: linear			
MOLECULE TYPE: DNA (genomic)			
FEATURE:			
NAME/KEY: CDS			
LOCATION: 350..14002			
FEATURE:			
NAME/KEY: CDS			
LOCATION: 14046..20036			
FEATURE:			
NAME/KEY: CDS			
LOCATION: 20110..31284			
FEATURE:			
NAME/KEY: CDS			
LOCATION: 31329..36071			
FEATURE:			
NAME/KEY: CDS			
LOCATION: 36155..41830			
US-08-804-198-1			
Query Match			
Best Local Similarity 9.5%; Score 109.8; DB 2; Length 44377;			
Matches 514; Conservative 0; Mismatches 607; Indels 10; Gaps 3			

Query Match	9.58;	Score 109.8;	DB 2;	Length 44377;
Best Local Similarity	45.48;	Pred. No. 5.2e-09;		
Matches 514; Conservative	0;	Mismatches 607;	Indels 10;	Gaps 3;

XX New claimed DNA (A) has one of 3 isolated *Mycobacterium*
CC tuberculosis DNA sequences of 3946 bp (I) (T93610), 3653 bp (VI)
CC (see T93611) or 440 bp (IX) (see T93612), or hybridises to, or is a
CC fragment of these sequences. Also claimed are RNA transcripts
CC of (A), and proteins encoded by (A). To isolate (I), a cosmid
CC library of *M. tuberculosis* DNA was screened with degenerate probes
CC (see T93621) containing GC-rich regions and designed to encode part
CC of a protein-rich protein. (I) contains open reading frames (ORFs)
CC for 4 proteins (see W31852-55). These ORFs can be amplified and
CC cloned into vectors for expression in *E. coli* or other hosts.
CC Clones (VI) and (IX) were obtained from a phage library and
CC together encode 3 proteins (see W31851 and W31856-57). The claimed
CC DNA sequences can be used for diagnosing tuberculosis and other
CC mycobacterial infections in humans or animals, for identifying
CC mycobacteria in (clinical) samples by hybridisation or
CC amplification, including differentiation between strains, as well
CC as for epidemiological studies, for monitoring vaccination or for
CC the development of anti-mycobacterial drugs and vaccines.

Sequence 3946 BP; 518 A; 1541 C; 1345 G; 542 T; 0 other;

Query Match	29.2%;	Score 338.8;	DB 18;	Length 3946;
Best Local Similarity	60.2%;	Pred. No. 8.6e-35;		
Matches 719; Conservative	0;	Mismatches 412;	Indels 63;	Gaps 7

[illegible][illegible]

RESULT	2	
ID	V64557	
ID	V64557 standard; DNA; 2367 BP.	
AC	V64557;	
DT	27-JAN-1999	(first entry)
DE	M. tuberculosis immunogenic polypeptide RD1E7 DNA.	
KW	Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB	
KW	vaccine; pharmaceutical; infection; diagnosis; ss.	
OS	Mycobacterium tuberculosis.	
PN	W09816646-A2.	
XX	23-APR-1998.	
XX	07-OCT-1997;	97WO-US18293.
XX	13-MAR-1997;	97US-0818112.
XX	11-OCT-1996;	96US-0730510.
XX	(COR1-) CORIXA CORP.	
PI	Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;	
PI	Reed SG, Skeiky YAM, Twardzik DR, Vedvick TS;	
DR	WPI; 1998-261042/23.	
DR	P-PSDB; W81744.	
PT	Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used	
PT	to develop products for the detection of M. tuberculosis infection	
PT	and for diagnosis, treatment and prevention of tuberculosis	
PS	Claim 4; Page 194-195; 230pp. English.	


```
QY 129 gccgcatcattgacgccccgcgcgtgcacacgcagccgcatgacgcgcattg 188
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 582 gtgcggcgcttggctgtacggccgacggccggttgccgcttgccggccgacac 523
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 189 gtccgacttgagcgccgcaacctccggtgccccgcgcgcgcgcgcagcaatgc 248
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 522 ggcggccacccgtacacggccgaccccgccggtgtgcgcgagacgtgcgacaccg 466
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 249 gccggagacgcgcgcatgcccgcgcagcgtcctcttgcgcgcgtgcccgcgcgc 308
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 465 gccgtcccccgcgtgcgcgcgtggtgcgcgcgcgcgcgcgcgcgcgcgcgcgc 406
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 309 gccgagacgcgttcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 368
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 405 gccggccgcttcgcccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 346
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 369 gccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 428
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 345 cagggccgagacccgtggcccccattgcgcgcgcgcgcgcgcgcgcgcgcgcgc 286
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 429 gccggtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 488
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 285 gccgtgacacacacacacacacacacacacacacacacacacacacacacacac 226
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 489 gtccctccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 548
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 225 ggcggccgcttcgcccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 166
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 549 gaacagacccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 608
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 165 ggtggg---gtgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 109
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 609 cccgcgcgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 668
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 108 accgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 49
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 669 ggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 716
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 48 tttctccgcccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
PT and diagnosis of tuberculosis
XX
PS Claim 4; Page 187-188; 250pp; English.
XX
CC This is the 3' region of DNA coding for an antigenic portion of
CC Mycobacterium tuberculosis antigen XP25; 5' DNA is provided in
CC V44438. XP25 DNA was isolated from a M. tuberculosis strain Erdman
CC genomic DNA expression library using sera from patients having
CC extrapulmonary tuberculosis. It bears no similarity to known
CC sequences. The invention relates to methods and compositions for
CC diagnosing tuberculosis. It provides polypeptides (see
CC W64291-W64379) comprising an antigenic portion of a soluble M.
CC tuberculosis antigen, or an immunogenic portion of a M.
CC tuberculosis antigen, as well as DNA sequences encoding such
CC polypeptides, recombinant expression vectors and transformed or
CC transinfected host cells. Also claimed are methods and diagnostic
CC kits for detecting M. tuberculosis infection in a patient using
CC these polypeptides, antibodies or oligonucleotide probes and
CC primers.
XX
SQ Sequence 985 BP; 138 A; 330 C; 423 G; 94 T; 0 other:

Query Match 18.5%; Score 215.2; DB 19; Length 985;
Best Local Similarity 58.5%; Pred. No.1.6e-19;
Matches 414; Conservative 0; Mismatches 288; Indels 6; Gaps 2:

QY 9 gccagtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 68
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 702 gcccaaacccgcttgccgcttgccgacacacacacacacacacacacacacacacac 643
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 69 gccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 128
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 642 gtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 563
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 129 gccgcatcattgacgccccgcgcgtgcacacgcagccgcatgacgcgcattg 188
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 582 gtgcggcgcttggctgtacggccgacggccggttgccgcttgccggccgacac 523
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 189 gtccgacttgagcgccgcaacctccggtgccccgcgcgcgcgcgcagcaatgc 248
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 522 ggcggccacccgtacacggccgaccccgccggtgtgcgcgagacgtgcgacaccg 466
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 249 gccggagacgcgcgcatgcccgcgcagcgtcctcttgcgcgcgtgcccgcgcgc 308
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 465 gccgtcccccgcgtgcgcgcgtggtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 406
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 309 gccgagacgcgttcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 368
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 405 gccggccgcttcgcccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 346
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 369 gccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 428
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 345 cagggccgagacccgtggcccccattgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 286
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 429 gccggtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 488
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 285 gccgtgacacacacacacacacacacacacacacacacacacacacacacacac 226
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 489 gtccctccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 548
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 225 ggcggccgcttcgcccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 166
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 549 gaacagacccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 608
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 165 ggtggg---gtgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 109
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 609 cccgcgcgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 668
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 108 accgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 49
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 669 ggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 716
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 12, 2001, 20:08:02 ; Search time 6876.06 Seconds
(without alignments)
864.115 Million cell updates/sec

Title: US-09-461-774-9
Perfect score: 1161
Sequence: 1 aatcgatgccagtcgccgc.....cgccgcagtaagtttga 1161

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1118133 seqs, 2558875100 residues
tal number of hits satisfying chosen parameters: 2236266

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*

1: gb_bal:*
2: gb_ba2:*
3: gb_om:*
4: gb_ov:*
5: gb_ph:*
6: gb_pl1:*
7: gb_pl2:*
8: gb_pr1:*
9: gb_pr2:*
10: gb_pr3:*
11: gb_ro:*
12: gb_sy:*
13: gb_un:*
14: em_fun:*
15: em_hum1:*
16: em_hum2:*
17: em_in:*
18: em_om:*
19: em_or:*
20: em_ov:*
21: em_pat:*
22: em_ph:*
23: em_pl:*
24: em_ro:*
25: em_sts:*
26: em_sy:*
27: em_un:*
28: em_vi:*
29: gb_htg1:*
30: gb_htg2:*
31: gb_in1:*
32: gb_in2:*
33: em_bal:*
34: em_ba2:*
35: em_hum3:*
36: em_hum4:*
37: gb_pr4:*
38: gb_htg3:*
39: gb_htg4:*
40: gb_htg5:*
41: gb_htg6:*
42: gb_htg7:*
43: em_htg1:*

44: em_htg2:*
45: em_htg3:*
46: em_hum5:*
47: gb_pl3:*
48: gb_pr5:*
49: gb_htg8:*
50: gb_htg9:*
51: gb_htg10:*
52: gb_htg11:*
53: gb_htg12:*
54: gb_htg13:*
55: gb_htg14:*
56: gb_in3:*
57: gb_htg15:*
58: gb_htg16:*
59: gb_htg17:*
60: em_htg4:*
61: em_htg5:*
62: em_htg6:*
63: em_htg7:*
64: em_hum6:*
65: gb_htg18:*
66: gb_htg19:*
67: gb_htg20:*
68: gb_htg21:*
69: gb_htg22:*
70: gb_htg23:*
71: gb_v11:*
72: gb_v12:*
73: gb_ba3:*
74: em_htg8:*
75: em_htg9:*
76: em_htg10:*
77: gb_pr6:*
78: gb_pr7:*
79: gb_sts1:*
80: gb_sts2:*
81: gb_pat1:*
82: gb_pat2:*
83: em_htg0:*
84: gb_htg24:*
85: gb_pr8:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	1156.2	99.6	32514	2	MTCVY130
2	1045	90.0	37036	2	AD000018 Mycobacte
3	413.4	35.6	30850	2	MTCTA1A11
4	413.2	35.6	56414	2	MTV002
5	377.4	32.5	68848	2	MTV003
6	367.4	32.4	31300	2	MTV035
7	367.2	31.6	47852	2	MTV023
8	360.8	31.1	40790	2	MTCV493
9	347.8	30.0	38900	2	MTCT63
10	344.4	29.7	39150	2	MTCT91B4
11	343.6	29.6	29550	2	MTV019
12	341	29.4	67200	2	MTV017
13	340.6	29.3	37586	2	MTCT270
14	338.8	29.2	3946	1	AF071081
15	338.8	29.2	3946	81	A93410
16	338.4	29.1	28826	2	MTV041
17	336.2	29.0	68848	2	MTV043
18	332.6	28.6	40163	2	MTCT28
19	328.6	28.3	11700	2	MTCT1418B
20	328	28.3	53662	2	MTV016
21	326.4	28.1	40838	2	MTY25D10
21	326.4	28.1	40838	2	MTY25D10

AUTHORS
Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry III, C.E., Tekkala, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S., Hamlin, N., Holroyd, S., Hornsby, T., Jagsis, K., Krogh, A., McLean, J., Moule, S., Murphy, L., Oliver, S., Osborne, J., Quail, M.A., Rajandream, M.A., Rogers, J., Rutter, S., Seeger, K., Skelton, S., Squares, S., Squires, R., Sulston, J.E., Taylor, K., Whitehead, S. and Barrett, B.G.

TITLE
Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence

JOURNAL
Nature 393 (6685), 537-544 (1998)

MEDLINE
98295987

REMARK
Erratum: [[published erratum appears in Nature 1998 Nov 12;396(6707):190]]
2 (bases 1 to 30850)

REFERENCE
Parkhill, J.

AUTHORS
Direct Submission

JOURNAL
Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium tuberculosis sequencing and mapping teams: Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk
On Jun 27, 1998 this sequence version replaced gi:1483528.

COMMENT
Notes:
Details of M. tuberculosis sequencing at the Sanger Centre are available on the World Wide Web.
(URL, http://www.sanger.ac.uk/Projects/M_tuberculosis/) CDS have been renumbered from the original cosmid submissions but the old gene designations are in brackets after the new gene numbers.
Gene prediction was based on a Hidden Markov Model of TR genes implemented in TRparse (Krogh) supplemented with visual inspection of positional base preference in codons, especially where there is an increase in the observed/expected third position G + C.
CAUTION: In some cases we may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, or ttg) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.

FEATURES
source
1..30850
/organism="Mycobacterium tuberculosis"
/strain="H37Rv"
/db_xref="taxon:1773"
/clone="Y1A11"
169..1071
/gene="Rv1814"
169..1071
/gene="Rv1814"
/note="Rv1814", (MTCY1A11.26c), unknown, len: 300, weak similarity to eg ERG3_YEAST p3235 c-5 sterol desaturase (365 aa), FASTA scores, opt: 154, z-score: 191.4, E(): 0.0011, (22.9% identity in 288 aa overlap)"
/codon_start=1
/transl_table=1
/product="hypothetical protein Rv1814"
/protein_id="CAB01464.1"
/db_xref="GI:1483557"
/db_xref="SWISS-PROT:Q50619"
/translation="MRDVLFAIPCFLLLILEWTARKLESTETATGQPRPASGAYLTRDSVASISMGVLSIATTAQMSKSLALGAAIYAYLAPMLSHRWYTVWIAVGD LLYSYVRIARHRAPLIWTQAHSSSEYFNPAATLROKNNSCGIIIMVPLPMGIRP WMYFCSWLNLIITQFWYTERIDRLPRFEFVNTPSHHRYHGMDEVYLDKNTGCLT IIMDRLEGSFQPELFRPHYGLTRKVDFTNWKLTQTRVVAIVDRWSATRLRLRLGIV FGPPGMEPRTRIDSNAAASLVTSR"

gene
1176..1841
/gene="Rv1815"
1176..1841
/gene="Rv1815"
/note="Rv1815", (MTCY1A11.28c), unknown, len: 221, similar to hypothetical protein G473456 M. fortuitum (255 aa), FASTA scores, opt: 182, z-score: 218.6, E(): 3.2e-05,

(29.6% identity in 230 aa overlap)"
/codon_start=1
/transl_table=1
/product="hypothetical protein Rv1815"
/protein_id="CAB01480.1"
/db_xref="GI:3261626"
/translation="MVRVLRARFAVALLAGSPATASADPVLVPRGMEIRQDNIV CTIGVYPALAKIAFTYAGHCRGGCAVTSRDYKVIQHLAIADNTPSGSTVATHELADY EAIVLADVDVTSNLTSPGRALSPGVVLPFGQAVCHFEVSTGECTGVSVNNGFT MSHGVLSSEKDSGGSPVYLPADGGPAQIVGIFNSWGGCPAAVSWRSTEQVHADLGVT PLA"

RBS
1893..1899
/note="gatgac, possible rbs upstream of Rv1816"
1904..2608
/gene="Rv1816"
1904..2608
/gene="Rv1816"
/note="Rv1816", (MTCY1A11.27c), len: 234, unknown, possibly involved in regulation, contains helix-turn-helix motif at aa 38-59, (+4.30 SD)."
/codon_start=1
/transl_table=1
/product="hypothetical protein Rv1816"
/protein_id="CAB01463.1"
/db_xref="GI:1483555"
/db_xref="SWISS-PROT:Q50617"
/translation="MCOTCRVGRKRRDREQIEAKIVYLGRLDHCAGLSIRAIR NLGAVSSAVYRVSSRDELTLTLVDVYSLQADLVDRDQVADSDVIAIARV RGMVATPBARMALVIGSPVGVNAPDRTAGVATRVVGAFFDAIAGIATGDIRLTD VAPDPSDEKITEQEPFGDDRVYTKCEFLMAGVGAISLVEFGYGDAMLTDPGV VPDQTRLVAIVAEH"

gene
3243..4706
/gene="Rv1817"
3243..4706
/gene="Rv1817"
/note="Rv1817", (MTCY1A11.26c), flavoprotein similar to eg G746486 flavoprotein subunit of fumarate reductase fad domain homologue (474 aa), FASTA scores, opt: 223, z-score: 250.0, E(): 5.7e-07, (24.1% identity in 489 aa overlap)"
/codon_start=1
/transl_table=1
/product="hypothetical protein Rv1817"
/protein_id="CAB01462.1"
/db_xref="GI:1483554"
/db_xref="SPRMBL:Q50616"
/translation="MSTYDIPATVSAETVTSMSDDVDVTVIGFIAGCAVAASAAAGA RVLVLERAAAGGTTALAGGFYLGCTVQOLATGHDSPEEMKYLVASREPDHPK IRAYCDGSVEHPNMLEGLGOFERSYPRGAIVIQPNDEGLMFTGNEKVMFLELAVPA PRGKRVYVPGDTGAANVIDLLKRAASLGIQIRFETGATBELIYDGTGKTVGMKRE SETGAIRAKSVYIIAAGFVMPKPDVAKVTRFLAEKPEVLCNTTDDGIGRLGVASAGA TOHNDQMFITAPRYPSPSILITGLIVNVLGQFEVAEDYHSRTAGFIQEPDSAAVYLIV DEAHLEHKKMPLVPLIDGMEVMEALGIPGNIATLDRYAAVARADDPFHQO PEFLLAODNPGMCAFDMSLGRAMVAGTIGLATISVYGQVLRDQGAVVAGIYAVGACA SNIAODKGYASGCTOLGSGSFEFGRAAHAARQA"

gene
complement(4826..6322)
/gene="PE_PGRS"
complement(4826..6322)
/gene="PE_PGRS"
/note="Rv1818c", (MTCY1A11.25), len: 498, ; Member of PE-family, glycine rich PGRS-subfamily, contains 2 x P500583 pKb family of carbohydrate kinases signature 1"
/codon_start=1
/transl_table=1
/product="PE_PGRS"
/protein_id="CAB01461.1"
/db_xref="GI:1483553"
/db_xref="SWISS-PROT:Q50615"
/translation="MSFVYTIPELAAVAVTDIAGIGSTICTANAAAAPPTTYUAAA DEVSAAMALFSGHQAQVQALSAQALFHQEFRAIAGAGSTAAABAASAAAPLEGVL DVINAPALALLGRPLIGNGANGAPGTGANGDGIILGNGAGSGSAGMPGNGGAGN GLFNGGAGGAGGAVASGTAGFGAGGAGGLIYGAGGAGGAGGAGGAGGAGGAGGAG GNGGGLIFGAGGAGGAGVGLAADAAGDAGDGLIFPGVGAAGAGGAGGTIVTGGAGGA

TITLE
JOURNAL
MEDLINE
REMARK
REFERENCE
AUTHORS
TITLE
JOURNAL

Hamlin,N., Holroyd,S., Hornsby,T., Jagels,K., Krogh,A., Mclean,J.,
Moule,S., Murphy,L., Oliver,S., Osborne,J., Quail,M.A.,
Squardream,M.A., Rogers,J., Rutter,S., Seeger,K., Skelton,S.,
Squires,S., Squires,R., Sulston,J.E., Taylor,K., Whitehead,S. and
Barrell,B.G.
Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence
Nature 393 (6685), 537-544 (1998)
98295987
Erratum: [[published erratum appears in Nature 1998 Nov
12:396(6707):190]]
2 (bases 1 to 56414)
Parkhill,J.
Direct Submission
Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium
tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome
Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique
Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux,
75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk
On Jun 27, 1998 this sequence version replaced gi:2624256.
Notes:
Details of M. tuberculosis sequencing at the Sanger Centre are
available on the World Wide Web.
(URL, <http://www.sanger.ac.uk/Projects/M.tuberculosis/>) CDS have
been renumbered from the original cosmid submissions but the old
gene designations are in brackets after the new gene numbers.
Gene prediction was based on a Hidden Markov Model of TB genes
implemented in TParse (Krogh) supplemented with visual inspection
of positional base preference in codons, especially where there is
an increase in the observed/expected third position G + C.
CAUTION: In some cases we may not have predicted the correct
initiation codon. Where possible we choose an initiation codon
(atg, gtg, or ttg) which is preceded by an upstream ribosome
binding site sequence (optimally 5-13bp before the initiation
codon). If this cannot be identified we choose the most upstream
initiation codon.

Location/Qualifiers
1. 56414
/organism="Mycobacterium tuberculosis"
/strain="H37Rv"
/db_xref="taxon:1773"
1. 233
/organism="Mycobacterium tuberculosis"
/strain="H37Rv"
/db_xref="taxon:1773"
/clone="Y154"
1. 56414
/note="fragment designated v002. Does not represent a
physical clone"
complement(3. 527)
/gene="recX"
complement(3. 527)
/note="Rv2736c, (MTV002.01c), len: 174 aa, recX, similar
to eg. RECX_PSEAE_P37860 regulatory protein recX from
Pseudomonas aeruginosa (153 aa), fasta scores; opt: 161
z-score: 257.2 E(): 3.6e-07, 30.7% identity in 137 aa
overlap. Overlaps and extends CDS from overlapping cosmid
MTCY154.16c"
/codon_start=1
/transl_table=11
/product="recX"
/protein_id="CA15532.1"
/db_xref="GI:2624257"
/db_xref="SPTREMBL:O33280"
/translation="MTVSCPPTSSESEEDARALCLTLTARSRTRAELAGOLAKRCY
PEDIGNAGVDRILAAGVLDPTDFAEGVQVOSRRARAASKRRLAEHLAKGVDDVITTT
VLGIDAGAEGRGAEKLVRRRLRREVLIDGTDEARVSRRLVAMLARHGIGTGLACEV
VLAELAEERRRR"
complement(493. 2865)
/gene="recA"
complement(493. 2865)
/gene="recA"

misc_feature

misc_feature

misc_feature

misc_feature

RBS

gene

CDS

gene

CDS

/note="Rv2737c, (MTV002.02c), len: 790 aa, recA,
identical to RECA_MYCTU_P26345 recA protein (790 aa).
contains self-splicing protein element (intron) from 2294
to 969 (c), similar to Intron II from TR:E332317
(EMBL:Y13030) DNA-directed DNA polymerase (EC 2.7.7.7) from
Thermococcus sp. (1829 aa), fasta scores; opt: 81 z-score:
235.2 E(): 6e-06, 24.6% identity in 183 aa overlap.
Contains PS00017 ATP/GMP-binding site motif A (p-loop),
PS00321 recA signature, and PS00861 Protein splicing
signature. See Davis et al, (1992) Cell 71(2):201-210"
/codon_start=1
/transl_table=11
/product="recA"
/protein_id="CA15533.1"
/db_xref="GI:2624258"
/db_xref="SWISS-PROT:P26345"
/translation="MTQTPREKALDELAVQIEKSYGKGYMRIGDEARQPISVTPG
SLDVALVIGIGLGRGVIEIYGPSSGKTTVALHVAANQAAGVAFDAEIALDP
DYAKKLVDDSLVSDPDGEOLAEIDMLIRSGALDIYVDSVAALVPRALEEGM
GDSHVGIOARLMSQALRKMGALNNSGTTAFITNQLDPKIGVMGSEPTTGKALKE
YASVMDYRREYETLKDGTMAGNTRKRVKVNKKCLAGCTRIFDPVETTHRIEDVVG
RKPIHVVAANADGTLHARPVSWFDQGTROVYIGRIAGGAIWATPDKVLTGEMRA
AGELRKGDVRAQPRRFDGSDAPIPADHARLGLYLIGDGDGNGGKTPINFTNVOR
ALIDVTRIAATLCAAHPOGRISLAIAHGRGENGVADICCOAGYIGKLAKEETIN
WFEPPDIAADIVGNLIRGLFESDGVSRBDGTALRGVYTTTSEQLAHQIHLLRFGV
GSTYRDVDPQKRSTYNGRIQGRQVFEYRISGMNVTARFASVYMKMPRGALIQ
AIPATGRRKGSQATYTLAEMTDVAVYLDERKVTQAEAAAGVAGSDGRGMKQV
LGASRLRDRVQALADALDDKFLHMDLAELRSVIEVLPTRARFDEVEGLTL
VABEVVHNSPPKQAEFLDYLGKISRELSLIDMGVDGGLIRKSGAMFTYEEEOIG
QGRKANRFEVENADVDEIEKLIKELGIGAVYTDPSPNDGVLPAVD"
complement(790. 807)
/gene="recA"
/note="PS00861 Protein splicing signature"
complement(790. 2112)
/gene="recA"
/note="recA Intein"
complement(2197. 2223)
/gene="recA"
/note="PS00321 recA signature"
complement(2644. 2667)
/note="PS00017 ATP/GMP-binding site motif A (p-loop)"
complement(12873. 2877)
/note="possible RBS upstream of Rv2737c"
complement(3247. 3453)
/gene="Rv2738c"
complement(3247. 3453)
/gene="Rv2738c"
/note="Rv2738c, (MTV002.03c), len: 68 aa, Unknown,
N-terminus is highly similar to the N-terminus of the
upstream ORFMTV002.07c (78.4% identity in 37 aa overlap);
also similar to AL020958|SC4H8.5 Streptomyces coelicolor
cosmid 4H8 (64aa) opt: 185 z-score: 283.5 E(): 2.9e-08;
39.7% identity in 63 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv2738c"
/protein_id="CA15534.1"
/db_xref="GI:2624260"
/db_xref="SPTREMBL:O33281"
/translation="MLAGVRLTEHERVALHFGAAYSSVLLDHLVLTGFGRSAAQAI
EDGVEPRDVMRALCADFDVPHDRM"
complement(3464. 4630)
/gene="Rv2739c"
complement(3464. 4630)
/note="Rv2739c, (MTV002.04c), len: 388 aa, Probable
transferase, similar to eg. TR:Q51560 (EMBL:L28170)
Pseudomonas aeruginosa thiamosyl transferase (426 aa),
fasta scores; opt: 178 z-score: 226.3 E(): 1.9e-05, 25.9%
identity in 425 aa overlap. Equivalent to Mycobacterium
leprae protein MLCB33_02c (392 aa); fasta scores
gp1294723|MLCB33_2 Mycobacterium leprae cosmid B33 opt:

TITLE	Barrell,B.G. Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence Nature 393 (6685), 537-544 (1998)	CDS	complement(1211..1393) /gene="Rv0810c" /note="Rv0810c" (MTV043.02c), len: 60. Unknown. Tbpase score is 0.876" /codon_start=1 /transl_table=11 /product="hypothetical protein Rv0810c" /protein_id="CA17616.1" /db_xref="GI:2916868" /db_xref="SPRMBL:053824" /translation="MGRGAKAKQTKVARELYSSPQTFDFLQRELSGTGTDRDGDGPSDDSDMNBEDWRR" complement(1401..1406) /note="possible RBS for Rv0810c" complement(1540..2646) /gene="Rv0811c" complement(1540..2646) /gene="Rv0811c" (MTV043.03c), len: 368. Unknown but equivalent to Mycobacterium lepreae hypothetical protein U2266F (366 aa). FASTA scores: U15182 MLU15182_13 (366 aa) opt: 1870 z-score: 2265.5 E(): 0. 77.4% identity in 367 aa overlap. Tbpase score is 0.912" /codon_start=1 /transl_table=11 /product="hypothetical protein Rv0811c" /protein_id="CA17617.1" /db_xref="GI:2916869" /db_xref="SPRMBL:053825" /translation="MAAVPADPDPDAGAIWHYGDPLGEORAGADAVLDRSHRAVLTLDGDRQTLHSTISQHVSDLPDAGSTONSLDGGQREVEDHVIQTEGGTYVDTPEWGRBRLAYLRKVMFMSMTVPRADMAVLSLGRRLAEFRYDALGLDVLPAELAVPLAGGCIYRRMPDGLAGCIELDVYKRDKDRDMORRLQAGYRPGIYAEVHRANHYPARRRGVSDTDEITIPHEVGWIGPGAGAVHLNKKGYRQDETVA RVHNIGRPRMLV LHLHDESVORPSTGDAVLVAGRTVGRGTVEHVEGLPVALALRLGLPGDTALVTGP EAEVAAYIDVDSLPPADVDGAGRAVRRLRGIR" 2716..2719 /note="possible RBS for Rv0812" 2729..3598 /gene="pabc" 2729..3598 /gene="pabc" 2729..3598 /gene="pabc" sp P283051 PABC_ECOLI_4-AMINO-4-DEOXYCHORSATE LYASE (269 aa). FASTA scores: opt: 165 z-score: 205.5 E(): 0.00064, (26.8% identity in 198 aa overlap). Also similar to several IIVE proteins. Tbpase score is 0.915" /codon_start=1 /transl_table=11 /product="pabc" /protein_id="CA17618.1" /db_xref="GI:2916870" /db_xref="SPRMBL:053826" /translation="MVTYLGSEILQPGMPLIHLADDLAAVRGDGVFETLVRGRACLYEAHILQRTQSARLMDLPEPDLPRRRRAVEVATQWVASTADEGALRTILYSGRGSA PTAIVKVSFVPAVITGARBDQVSAITLIDRLPAGCGAMPLISATLTSYRMAVAL RHARQAGADVITFEVTDGYVLEGRSVTVLATDDGGGNCCLTLPWPVPIILNGTQ QALEVRAARNGYDCDYLALRYADLDFSQIWLVSMTLAARVHTLIDGRRLRKPPIAEVFAELVDAAYSDR" complement(3644..4324) /gene="Rv0813c" complement(3644..4324) /gene="Rv0813c" (MTV043.05c), len: 226. Unknown but highly similar to M. lepreae cosmid B2269P U15182 MLU15182_16 (242aa). FASTA scores: opt: 1191 z-score: 1405.0 E(): 0. 78.3% identity in 226 aa overlap. Also similar to mtcy056c_37, Hypothetical protein, 30.4% identity in 171 aa overlap. /codon_start=1 /transl_table=11 /product="hypothetical protein Rv0813c" /protein_id="CA17619.1"
REFERENCE	2 (bases 1 to 68848) Erratum: [[published erratum appears in Nature 1998 Nov 12:396(6707):190]]	RBS	
AUTHORS	Parkhill,J.	gene	
TITLE	Direct Submission	CDS	
JOURNAL	Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA unite de Genetique Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk On Jun 27, 1998 this sequence version replaced gi:2916866.	gene	
COMMENT	Notes: Details of M. tuberculosis sequencing at the Sanger Centre are available on the World Wide Web. (URL, http://www.sanger.ac.uk/Projects/M_tuberculosis/) CDS have been renumbered from the original cosmid submissions but the old gene designations are in brackets after the new gene numbers. Gene prediction was based on a Hidden Markov Model of TB genes implemented in Tbpase (Krogh) supplemented with visual inspection of positional base preference in codons, especially where there is an increase in the observed/expected third position G + C. CAUTION: In some cases we may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, or ttg) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.	misc_feature	
FEATURES	Location/Qualifiers	source	
source	1..68848 /organism="Mycobacterium tuberculosis" /strain="H37Rv" /db_xref="taxon:1773" 1..67906 /note="fragment designated v043. Does not represent a physical clone" 1..40 /organism="Mycobacterium tuberculosis" /strain="H37Rv" /db_xref="taxon:1773" /clone="Y7H7A" 31..1125 /gene="purm" 31..1125 /gene="purm"	source	
misc_feature	/note="Rv0809, (MTV043.01), len: 364. purM, probable 5'-phosphoribosyl-5-aminoimidazole synthetase, similar to many e.g. P12043 PUR5_BACSU (346 aa) and Salmonella typhimurium U68765 STU68765_2 (345 aa). FASTA scores: sp P12043 PUR5_BACSU PHOSPHORIBOSYLFORMYLGLYCINAMIDINE (346 aa) opt: 1023 z-score: 1250.3 E(): 0. 46.5% identity in 331 aa overlap. andU68765 STU68765_2 (345 aa) opt: 1014 z-score: 1239.4 E(): 0. 47.6% identity in 330 aa overlap. Tbpase score is 0.901" /codon_start=1 /transl_table=11 /product="purm" /protein_id="CA17615.1" /db_xref="GI:2916867" /db_xref="SPRMBL:053823" /translation="MTDLAKPGCKDPSRGITVASAGVDIEAGDRAIDLKPLASKAT RPEVGGIGGAGFAGFTLRAGDYREFVLAASSGSGVGTKLAIQAMKHDTGVGLVAMVY DDLYVCGAEPLFLDLYIAGRIIVERLSAIVAGIADCGMRACGALLGGEAEHGGLE PDHYDISATGAGVEADNVLGPDVRGVDVITIAMGSSGHSNGSIVAKYLETEDRMN LAGHVEEGRITGSELLEPRTIRYAKKDLALAAERTVTFCHVGGGAGNIDRIPIPG LIAEVDRTVPAVFTVIAORGRVARTEMKTFNMGVMAVAVAPEDTTRALAVLTA RHLDGVLTGVCCKGQGPRAKLVGQHRP" complement(1211..1393) /gene="Rv0810c"	gene	

[illegible][illegible]

MEDLINE 98295987
REMARK Erratum:[published erratum appears in Nature 1998 Nov
12;396(6707):1901]
REFERENCE 2 (bases 1 to 31300)
AUTHORS Parkhill,J.
TITLE Direct Submission
JOURNAL Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium
tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome
Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique
Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux,
75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk
On Jun 27, 1998 this sequence version replaced gi:2909461.
Notes:
Details of M. tuberculosis sequencing at the Sanger Centre are
available on the World Wide Web.
(URL, <http://www.sanger.ac.uk/projects/M-tuberculosis/>) CDS have
been renumbered from the original cosmid submissions but the old
gene designations are in brackets after the new gene numbers.
Gene prediction was based on a Hidden Markov Model of TB genes
implemented in TBparse (Krogh) supplemented with visual inspection
of positional base preference in codons, especially where there is
an increase in the observed/expected third position G + C.
CAUTION: In some cases we may not have predicted the correct
initiation codon. Where possible we choose an initiation codon
(atg, gtg, or ttg) which is preceded by an upstream ribosome
binding site sequence (optimally 5-13bp before the initiation
codon). If this cannot be identified we choose the most upstream
initiation codon.
Location/Qualifiers
source 1..31300
/organism="Mycobacterium tuberculosis"
/strain="H37Rv"
/db_xref="taxon:1773"
misc_feature <1..30489
/note="fragment designated v035. Does not represent a
physical clone"
RBS /note="fragment (331..335)
complement(331..335)
409..412 /note="possible RBS, ggaag, for Rv0273c"
RBS /note="possible RBS, ggaag, for Rv0274"
420..1001
/gene="Rv0274"
420..1001
/gene="Rv0274"
CDS /note="Rv0274" (MTV035.02), len: 193. Unknown, but some
similarity to gpIAJ002571BSA2571_7 Bacillus subtilis 168
56kb DNA (316 aa), opt: 138 z-score: 174.8 E(): 0.033;
27.1% identity in 133 aa overlap. Previous hits with
FOSFOWCIN-RESISTANCE PROTEIN from SERRATIA MARCESCENS.
Contains P50082 Extradiol ring-cleavage dioxygenases
signature near C-terminus. FASTA scores: TR:056415
(EMBL:M85195) FOSFOWCIN-RESISTANCE PROTEIN (141 aa), opt:
82 z-score: 285.5 E(): 1.1e-08, 29.1% identity in 151 aa
overlap. TBparse score is 0.914"
/codon_start=1
/transl_table=1
/product="hypothetical protein Rv0274"
/protein_id="CAAI7349.1"
/db_xref="GI:2909463"
/db_xref="SPTREMBL:O53680"
/translation="MIKPHNTEFELGHNVALGCSDMARTVDFYSNITGLPLIK
LDLPGGOGHGFEPAGNGDCVAFEPADAPRVGLSSPAVLPISGIDTSVSNMHL
AFHYPARERFDARLRKLDKGVRCVPVPLNHDDSEFQVSAVHPGVYVSFFYQDPDGI
LEFACWTKREFTSDAQAVPKTAADRRPPVADR"
misc_feature 840..905
/gene="Rv0274"
/note="P50082 Extradiol ring-cleavage dioxygenases
signature"
complement(931..1656)
/gene="fadD27"
complement(931..1656)
CDS /gene="fadD27"
/note="Rv0275c, (MTV035.03c), len: 241. fadD27, similar

toop|AL022004|MTV043.17 Mycobacterium tuberculosis seque
(213aa), FASTA scores: opt: 230 z-score: 266.0 E():
2.7e-07; 32.6% identity in 190 aa overlap. TBparse score is
0.896"
/codon_start=1
/transl_table=1
/product="fadD27"
/protein_id="CAAI7350.1"
/db_xref="GI:2909464"
/db_xref="SPTREMBL:O53681"
/translation="MTRSDRYRCVEAERLATRRRSAGLDLGSDDQNDVETL
RTICRRAGLSYRYESETDEKDEGVAFPMVVAEATVTAQAAVAPAREQV IAGNA
NIVRTITADAVGRFLRSTOLANAVITRKRESALFAMISGQAVPTLHAAPADHAK
AVAHFAGVCQCOTISAMLADVRIDPPQVLDQALALLDELTPDLSPKRVAAIAKSG
ANDPPEVACQPPSSARPARRS"
1734..1738
/note="possible RBS, ggaag, for Rv0276"
1746..2666
/gene="Rv0276"
1746..2666
/gene="Rv0276"
/note="Rv0276" (MTV035.04), len: 306. Unknown but similar
to Mycobacterium tuberculosis MTCY427.18 (296 aa). FASTA
scores: 270692|MTCY427.18 Mycobacterium tuberculosis
cosmid (296 aa) opt: 874 z-score: 1042.8 E(): 0;
49.6% identity in 282 aa overlap. TBparse score is 0.894"
/codon_start=1
/transl_table=1
/product="hypothetical protein Rv0276"
/protein_id="CAAI7351.1"
/db_xref="GI:2909465"
/db_xref="SPTREMBL:O53682"
/translation="MAISLVAHQPIPIVERPMADPPRLQLARRRSAGRCINEDSLM
GVALLAPANVIMELAMPGVGYGLSEKVESGRLRHP IKAARTFTYVAVVAGSD
QKAARRAVNVKHAQVSTEPSYSAFPEIQLMWACLYKGVVYRTPVCEMD
EADHHYRAGMAGTTLOVPPQMPDPPRAADRVYRQSLDVRHIDVVRVLYIVL
RIHGIALGPRLRSEGILITLITGFLPDRDRMRLPMWATKRRRDALMAVARTVN
RLMRYPREPPNMLMDLDRMRGRPLV"
complement(2706..3134)
/gene="Rv0277c"
complement(2706..3134)
/gene="Rv0277c"
/note="Rv0277c" (MTV035.05c), len: 152. Unknown but
similar to hypothetical Mycobacterium tuberculosis
proteins MTCY274.03 (147 aa) and MTV006.14 (143 aa). FASTA
scores: 274024|MTCY274.3(147 aa) opt: 357 z-score: 461.7
E(): 3.2e-18;41.4% identity in 140 aa overlap; and
AL02101|MTV006.14 (143aa) opt: 258 z-score: 337.6 E():
2.6e-11; 33.8% identity in142 aa overlap"
/codon_start=1
/transl_table=1
/product="hypothetical protein Rv0277c"
/protein_id="CAAI7352.1"
/db_xref="GI:2909466"
/db_xref="SPTREMBL:O53683"
/translation="MFLIDVNVLLAARHGHPNRTYRPWDRILLADDPPTVNVW
ASFRLITNRIFETSPRADAFVEVNAQPHLFTSPSPRH VILRLKDEADAS
GDLLPDAVLGAIVAHHCVAVSLDRDARFASVHRIPPI"
complement(3435..9071)
/gene="PE_PGRS"
complement(3435..6308)
/gene="PE_PGRS"
/note="Rv0278c" (MTV035.06c), len: 957. Member of M.
tuberculosis gly-, Ala-rich PGRS subfamily of PE protein
family. Similar to many others. Contains P500583 pfkB family
of carbohydrate kinases signature 1. FASTA scores:
2958901|MTCY28.25 (914 aa) opt: 3849 z-score: 2368.8 E():
0; 67.8% identity in 903 aa overlap. TBparse score is
0.870"
/codon_start=1
/transl_table=1
/product="PE_PGRS"
/protein_id="CAAI7353.1"

[illegible][illegible]

Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk
On Jun 24, 1998 this sequence version replaced gi:2131042.

COMMENT

Notes:
Details of M. tuberculosis sequencing at the Sanger Centre are available on the World Wide Web.
(URL, http://www.sanger.ac.uk/Projects/M_tuberculosis/) CDS have been renumbered from the original cosmid submissions but the old gene designations are in brackets after the new gene numbers.
Gene prediction was based on a Hidden Markov Model of TB genes implemented in TParse (Krogh) supplemented with visual inspection of positional base preference in codons, especially where there is an increase in the observed/expected third position G + C.
CAUTION: In some cases we may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, or ttg) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.

FEATURES

```
Source
    location/Qualifiers
        1..40790
        /organism="Mycobacterium tuberculosis"
        /strain="H37Rv"
        /db_xref="taxon:1773"
        <1..40722
        /organism="Mycobacterium tuberculosis"
        /strain="H37Rv"
        /db_xref="taxon:1773"
        /clone="Y493"
        <1..1638
        /organism="Mycobacterium tuberculosis"
        /strain="H37Rv"
        /db_xref="taxon:1773"
        /clone="Y21B4"
        complement(10..771)
        /gene="Rv1424c"
        complement(10..771)
        /gene="Rv1424c"
        /note="Rv1424c" (MTCY21B4.42c, MTCY493.30), len: 253.
        /note="Rv1424c" (MTCY21B4.42c, MTCY493.30), len: 253.
        Possible membrane protein, contains PS00402
        Binding-protein-dependent transport systems inner membrane
        comp signature.."
        /codon_start=1
        /transl_table=11
        /product="hypothetical protein Rv1424c"
        /protein_id="CAB09246.1"
        /db_xref="GI:2131072"
        /db_xref="SPTRMBL:P71693"
        /translation="MTVYVPGASPRSAVSRSYRCQVQASAGTSARSPSPRRP
        AEKLVFVPLGILTLTLASACOTASASGNEPRGIDRAILKLFVSMDLGMCINRFTYDS
        KLAAPSPQVACDSREARIRNDGFHANAPSCMRIDYELITGNHRAVYCLKLYLVGYC
        YPAVTPGKPSVLLVAPASCDSELPSPRVATALVPGRSANRERSRFVVEIKSLGA
        GRCDSASVSLQPEEIEGPAIPASSQLVCVARK"
        complement(520..606)
        /gene="Rv1424c"
        /note="PS00402 Binding-protein-dependent transport systems
        inner membrane comp signature"
        775..2154
        /gene="Rv1425"
        775..2154
        /gene="Rv1425"
        /note="Rv1425" (MTCY21B4.43, MTCY493.29c), len: 459.
        Similar to many M. tuberculosis hypothetical proteins e.g.
        MTCY28.26, fasta scores: E(): 0, 36.3% identity in 460 aa
        overlap"
        /codon_start=1
        /transl_table=11
        /product="hypothetical protein Rv1425"
        /protein_id="CAB09245.1"
        /db_xref="GI:2131071"
        /db_xref="SPTRMBL:P71694"
        /translation="MKRLSSVDAAFWASFTAGMHMHVGLALICDPSDAPEVSPQRLRE
```

```
LIIRLPEIIPOLRMWVTGAPLDRPMVEDEELIDFHIRIGRVPAPGGRLEELV
GRMSYKIDRSRPLMELMWTIEVGGRATLTKKHAIIVDSCAGCEILIDTPP
RPPQETWYFPGPIQIGERRATIALINGVTMPRTVRIILEYVHQOIALSVAGRP
ARITEAPKTRNPAPSPHRRVTGTREVLAKAKADFGVNLNVNVALVAGAROTL
OKRDELPAKPLIAQIPVSTRSEETKADGNVSSMTASLHIEDPAKRLAAHEROTL
SAKEMANAPSAHQIMGLTEPGLQLQAAAVASGISHMILAPINLVSVPQPPPP
LYMAGARLDSLVPLGPPMDVALNITCFSYQDYDFGLVTPPEVANDIDEMADAITEPA
LAELERRAE"
complement(2176..3438)
/gene="lipo"
complement(2176..3438)
/gene="lipo"
/note="Rv1426c" (MTCY493.28), len: 420. Unknown.
possibleesterase, similar to MTCY493.28, MTCY08D5.15,
MTCY21B4.10c highly similar esp. in C-terminus to M.
tuberculosis hypothetical protein Q50681 (431 aa) (50.1%
identity in 411 aa overlap), and also similar in central
region to AABD_HUMAN P22760 arylacetamide deacetylase (398
aa), fasta scores: opt:210 z-score: 248.8 E(): 7.6e-07,
29.3% identity in 191 aa overlap"
/codon_start=1
/transl_table=11
/product="lipo"
/protein_id="CAB09244.1"
/db_xref="GI:2131070"
/db_xref="SPTRMBL:P006832"
/translation="MRRFRRAKPPRLTAAVELLMANGRLPLSGSGTYLAWLWG
PTSEVPGVYLASVLDLRGRGRDFFGLKKAALALTAAMVILAVIRGATTPGP
VLEAGLIEQLGPDYAKELATLPTEPMRSGNLDLRLAMARRVETTNVVCYGR
ANLADILEMRDLPDPAKAPVLYVPGVAMVGMPPQAPPLMSHLAAGVCSLNR
VSPRHTMPDITVDVQRALAAWKENMTAAGPDPVVAISGSGAGLCLALATFNDPR
FQPGFEQDVTVAANAAPVYGRVDETTDAAPRRFEVGLTFEVVKKRSTRODIFVDA
SPIHVRADAPPPFVLGHRDLSLIPVAANAHFVEELRAVSSPAVYADLPHAQAFDV
FGSPRAHHTAAVAARFLSWYATNPPT"
complement(3438..5045)
/gene="fadD12"
complement(3438..5045)
/gene="fadD12"
/note="Rv1427c" (MTCY493.27), len: 535. fadD12, possible
long chain long-fatty-acid-coA ligase similar to eg.
PC60_YEAST P38137 peroxisomal-coenzyme a synthetase (543
aa), fastascores: opt: 507 z-score: 579.4 E(): 2.9e-25,
30.4% identity in 365 aa overlap Contains PS00455 Putative
AMP-bindingdomain signature. Also similar to many M.
tuberculosisproteins eg. MTC10644.14 (44.8% identity in 525
aa overlap)"
/codon_start=1
/transl_table=11
/product="fadD12"
/protein_id="CAB09257.1"
/db_xref="GI:2131069"
/db_xref="SPTRMBL:O06831"
/translation="MRTIQAQGLATMRAGLIAPLRPDRLRIIVAAARRBGMGTAG
FAGARBCPRDGRGLDELGTJLWROLDRGNALAAALDAPGPRVVGINCRNIRGF
VDALIAVNRICAHILINTSPFAGPALAEVYRVEDVYVDEFSATPRALAKPQA
TRIVAVDEDEDLIVYEKLVANAHARRPETHSHGKVIITLIGSTGTREGAHVSGGIG
TLKALIDTPWRABEVITVAPMFAWESOLVLAASSLACTIVTRRRDPKATIIDLID
RHHNTGLVVPVMEFDRIMDLPAEIRNRNDGSLRFLAASRMRDVLIAIMDQGDV
IYNNVNAIEAGMIATAPADLRTPADTGRRAEETETIILDOQETVEVGTGVTIYVR
NDSDQEDYTGSCAANDFAGFMSSGDVGLDNGRGLFYVGRDEMIVSGENITPREVE
KTIAHPVPAEAAVIGVDDOOGORLAFFVLKRGVSAPTPTLMQHVNDNLANKVPR
DIAYDELPRGITGKILTEQSRKRS"
complement(4422..4457)
/gene="fadD12"
/note="PS00455 Putative AMP-binding domain signature"
complement(5048..5050)
/note="possible RBS upstream of Rv1427c"
complement(5049..5876)
/gene="Rv1428c"
complement(5049..5876)
/gene="Rv1428c"
/note="Rv1428c" (MTCY493.26), len: 275. Unknown, some
similarity to YV29_MYCTU Q11167 hypothetical 39.3 kd
```


been renumbered from the original cosmid submissions but the old gene designations are in brackets after the new gene numbers. Gene prediction was based on a Hidden Markov Model of TB genes implemented in TParse (Krogh), supplemented with visual inspection of positional base preference in codons, especially where there is an increase in the observed/expected third position G + C. CAUTION: In some cases we may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, or ttg) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the Initiation codon). If this cannot be identified we choose the most upstream Initiation codon.

FEATURES

```
source
    1. 38900
        /organism="Mycobacterium tuberculosis"
        /strain="H37Rv"
        /db_xref="taxon:1773"
        <1. .37468
        /organism="Mycobacterium tuberculosis"
        /strain="H37Rv"
        /db_xref="taxon:1773"
        /clone="Y63"
        /note="possible RBS, GGAGC, for Rv0297"
        82. .1857
        /gene="PE_PGRS"
        82. .1857
        /gene="PE_PGRS"
        /note="Rv0297, (MTCY63.02), len: 591. Unknown Gly, Ala-rich protein belonging to the PGRS family (PE). Many relatives in M. tuberculosis. FASTA results: Y03A MCTU 010637 hypothetical glycine-rich 49.6 kd pro (603 aa) opt: 1884, E(): 0; 53.7% identity in 635 aa overlap. TParse score is 0.850"
        /codon_start=1
        /transl_table=11
        /product="PE_PGRS"
        /protein_id="CA809596.1"
        /db_xref="GI:2193918"
        /db_xref="SPTREMBL:007224"
        /translation="MSFVIAQPEMTIAAAGELASIRSAINMANAAAAAQTGVMSAA DEVSTAVALFSSHAQYQAASQAFAHQAQVRLTVDAQAYASAEANAGPNMLAA VNAPQALLGRPLTNGANGAPGTGAGGDDGLFNGNGSGSGAPGAGGAGAGF FGNNGNGDGGAGANGAGCTRAGPFPGCGNGAGCTIVACINGLGAGGDCGANGA FGNNGNGMGAGAGAGVNAVNPGLATPYTPANANGNLVGVPTAGGADGAGASA IQAGAGAGDGNASTSGIGIAQTGAGAGAGAGDAPGNGNGSGSEVHTGATGS SASGNGATGNGGAGGAGAGGNGGSHVSGSVNTAGAGKNGGTGGAGPGGHG SVLSPPVDSNGAGGAGGAGVSATDIAGTCGRCGNGHGLMTVNGDGGAGGVG VGAGAAAGATGGHGGGGSVNPPIGSGSEAGDGGKGLGDDGGRGIFPGFAGAGCA GAVGAGAGAGCTGGGCGGAGCAFFNACTPCAAGCTGCGVGTGAAGAGKAGSGGCVN CATGADGAKGLDGAATGKNGNPG"
        1988. .1991
        /note="possible RBS, AGGA, for Rv0298"
        2000. .2227
        /gene="Rv0298"
        2000. .2227
        /gene="Rv0298"
        /note="Rv0298, (MTCY63.03), len: 75. Function: unknown. TParse score is 0.891"
        /codon_start=1
        /transl_table=11
        /product="hypothetical protein Rv0298"
        /protein_id="CA809590.1"
        /db_xref="GI:2193919"
        /db_xref="SPTREMBL:007225"
        /translation="MTKEKISVTDAAVLAIDADARAALNRSEMIQALRNEHLRV ALRDYTAKEVPLDIDAYARVYQANRAAGS"
        2224. .2526
        /gene="Rv0299"
        2224. .2526
        /gene="Rv0299"
        /note="Rv0299, (MTCY63.04), len: 100. Function: unknown."
        /codon_start=1
```

```
/transl_table=11
/product="hypothetical protein Rv0299"
/protein_id="CA809597.1"
/db_xref="GI:2193920"
/db_xref="SPTREMBL:007226"
/translation="MIADGIAPRDSSEHLYVAVLSNALHRAADTGVITCPPIGR VPEDLAWVAANEQNGTILPELVQWLVHVALGAPLGNAGVAALREASVYATLAC"
2574. .2795
/gene="Rv0300"
2574. .2795
/gene="Rv0300"
/note="Rv0300, (MTCY63.05), len: 73. Function: unknown."
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv0300"
/protein_id="CA809591.1"
/db_xref="GI:2193921"
/db_xref="SPTREMBL:007227"
/translation="MSDVLIRIDIPDDVLASIDAIARLGLSTREYIRRLAQAQTAR VTVTADLRRLRGAAVAGLGDPELMQAMR"
2792. .3217
/gene="Rv0301"
2792. .3217
/gene="Rv0301"
/note="Rv0301, (MTCY63.06), len: 141. Function: unknown, has similarity to other hypothetical M. tuberculosis proteins. FASTA results: YW07_MCTU 010847 hypothetical 14.7 kd protein cy39.07c (132 aa) opt: 99; E(): 1; 28.1% identity in 135 aa overlap; MTCY159.10c (37.9% identity in 116 aa overlap), [MTV002_23.TParse score is 0.928"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv0301"
/protein_id="CA809598.1"
/db_xref="GI:2193922"
/db_xref="SPTREMBL:007228"
/translation="MTDORWLIDKSAIVRLTDSPPMETISNRIENGIVHTGVPLEV GFSACGEIARREFREPPLSAMPVEYLTPRIEDRLAEVOTLADGHHRGSPIDPLI AATVATLSGLTVLHVDKPDATLALTGQKTERLTHRPQA"
3338. .3341
/note="possible RBS, AAGC, for Rv0302"
3353. .3985
/gene="Rv0302"
3353. .3985
/gene="Rv0302"
/note="Rv0302, (MTCY63.07), len: 210. Function: unknown but similar to several repressors and regulatory proteins. FASTA results: ACRF_ECOLI P34000 potential acrab operon repressor (215 aa) opt: 172; E(): 3.1e-05; 22.7% identity in 194 aa overlap. TParse score is 0.896. Also similar in N-terminus to M. tuberculosis hypothetical regulator MTCY07A7.24 (38.7% identity in 62 aa overlap). Contains probable helix-turn helix motif from aa 36-57"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv0302"
/protein_id="CA809599.1"
/db_xref="GI:2193923"
/db_xref="SPTREMBL:007229"
/translation="MGVPAKKKQOQGESRESILDPATERLMTKGYATSI SDIRDAC GLAPSIYWHFSGKEGVLAAMERGAOFEPFAIPWDEHGVQDQSEKOLTEIVSIO SOHPDLFLFYLLSMERSODPAVAVVVRVRTATARRRDSITHLPLSPDIPGRKADLV VAEITFAVALSDGYEFAGHLEPDTTIDVERMYRRLQALDEALIVLEET"
3973. .3977
/gene="Rv0302"
3982. .4890
/note="possible RBS, GGAGC, for Rv0303"
3982. .4890
/gene="Rv0303"
3982. .4890
/gene="Rv0303"
/note="Rv0303, (MTCY63.08), len: 302. Function: unknown but similar to NADPH dehydrogenase and other NADPH oxidoreductases. FASTA results: Q42850 NADPH DEHYDROGENASE
```

[illegible][illegible]

of positional base preference in codons, especially where there is an increase in the observed/expected third position G + C.
CAUTION: In some cases we may not have predicted the correct initiation codon. Where possible we choose an initiation codon (aug, gcg, or tgg) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.

FEATURES

```
source
  location/Qualifiers
    1..39150
    /organism="Mycobacterium tuberculosis"
    /strain="H37Rv"
    /db_xref="taxon:1773"
    /clone="Y21B4"
    <1..721
    /organism="Mycobacterium tuberculosis"
    /strain="H37Rv"
    /db_xref="taxon:1773"
    /clone="Y2B12"
    689..693
    /note="possible RBS, GGGG, for Rv1385"
    703..1527
    /gene="pyrF"
    /gene="pyrF"
    703..1527
    /gene="pyrF"
    /note="Rv1385, (MTCY21B4_02), pyrF, len: 274, orotidine 5'-phosphate decarboxylase, identical to M. bovis, DCOF_MYCBOP42610 (274 aa); contains PS00156 Orotidine 5'-phosphate decarboxylase active site"
    /codon_start=1
    /transl_table=11
    /product="pyrF"
    /protein_id="CAB02190.1"
    /db_xref="GI:1542893"
    /db_xref="SWISS-PROT:P77898"
    /translation="MTGFGRLAEAKARGPCLGIDPHELLRGMDLATTADGIAAFCDICVRADEFAVVKPVQAFESYGAGFAVLERTIAELRADVLYLADAKGDIGATMSAYTAWGDSPLADAVTASPYLFGSLRPILLEAAGRGVFLAATSPBECAAVONAAADGRSVAGLVDOYGANEAANGPRGSGVYGATAPQAPPLSAFTGVIYPCVGVGGPRPALGGLGGAASSQLLPVAREVLRAGPVPPELRAGERMRDVAFLAAV"
    970..1011
    /gene="pyrF"
    /note="PS00156 Orotidine 5'-phosphate decarboxylase active site"
    1710..1713
    /note="possible RBS, GGAG, for Rv1386"
    1722..2030
    /gene="pe"
    1722..2030
    /gene="pe"
    /note="Rv1386, (MTCY21B4_03), len: 102. Member of M tuberculosis PE family, similar to many e.g. G913039 ORF 3 OF PCRS TANDEM REPEAT (polymorphic GC-rich sequence)(100 aa), fasta scores, opt: 149, E(): 0.0013, (31.5% identity in 92 aa overlap), also G699138 U1556A (99 aa) (34.7% identity in 95 aa overlap) and G466937 U1620K (100 aa) (36.2% identity in 69 aa overlap)"
    /codon_start=1
    /transl_table=11
    /product="pe"
    /protein_id="CAB02191.1"
    /db_xref="GI:1542894"
    /db_xref="SPTREMBL:P71656"
    /translation="WTLRVPESLAGASNAIEAVTARLAAHAHAAPPTAAVIPPQSDSVSYCAVEFVHSGQHVMAAQGVLELGRSGVGAESGASYAARDALAAAYLSGLG"
    2027..3646
    /gene="ppe"
    2027..3646
    /gene="ppe"
    /note="Rv1387, (MTCY21B4_04), len: 539. Member of M. tuberculosis ppe family of protein, similar to many e.g. Y05F_MYCTU Q10892 hypothetical 46.9 kd protein cy251.15
```

(463 aa), fasta scores: E(): 4.2e-26, 37.7% identity in 531 aa overlap; contains PS00343 Gram-positive cocci surface proteins 'anchoring' hexapeptide; similar also to MTCY274_23c (37.5% identity in 168 aa overlap)"

```
misc_feature
  /transl_table=11
  /codon_start=1
  /product="ppe"
  /protein_id="CAB02192.1"
  /db_xref="GI:1542895"
  /db_xref="SPTREMBL:P71657"
  /translation="MTEPWIAFPPEVSHAMNYGAGVPMILSATONGELSAQVAAASEVELLGVVASSEGWGOAAEFVAAYVFLAMLIQASDCVEMAAQOHVIEAYTAAVEIMPTOVLELANQIKLAVLVATNFNGINTPIPAIEVEYEMRAATVATYSTVSRSATLSAMPTSPSPILLIKSDLELPTDGTGSEGNHGGSHGHAINDPFAELRGVSAGRIYWDPNGLNLNDLDDYVYPGHAIWMLARGLEFPDQGFPELLFTPTNGAFQFLIVVVDLPHTIAQIATWLGQYFQULSALGLTAHIGATITGLGLSLATPSAAIPAVVPELTPVAAPMLAVAGVPAVAGMIPASAPAPAAGATAGPTTPPATGFGGFPYLVGGGPGIGFGSGSAHAAASDSAAASAAQASARAQARARGRSAAKARGHDEFEVTMDGFPDAAPAPRHOGARARSDCAGPAGTVRKEAVVKAAGLFTLAGDDFGGGPTMPMPGTWHDQVDFDEHR"
  2573..2590
  /gene="ppe"
  /note="PS00343 Gram-positive cocci surface proteins 'anchoring' hexapeptide"
  3952..4524
  /gene="mhf"
  3952..4524
  /gene="mhf"
  /note="Rv1388, (MTCY21B4_05), len: 190. Almost identical to, but longer than M. smegmatis PR:p96802 (EMBL:U75344) integration host factor (mhf) for mycobacteriophage U5 (105 aa), fasta scores: E(): 0, 96.1% identity in 102 aa overlap"
  /codon_start=1
  /transl_table=11
  /product="mhf"
  /protein_id="CAB02193.1"
  /db_xref="GI:1542896"
  /db_xref="SPTREMBL:P71658"
  /translation="MLGNTIHVPCQPCRHGAGPSRLGRPRADRPVAVATPYTLHVCFONQGVADPFVAKPEYGRIRMPVAPAGNNRLLISMKGIVALLPOLUDEQKAILEKAAARARAELKDLRLKRGSTNLTVLKAESPEVYLGKKKVSALLLEALPRKGVKAOEIMTELEIAPTRRLRGIDRORALLKFPESA"
  4649..4653
  /note="possible RBS, AGGAG, for Rv1389"
  4659..5285
  /gene="gmk"
  4659..5285
  /gene="gmk"
  /note="Rv1389, (MTCY21B4_06), gmk, len: 208. Probable guanylate kinase, similar to e.g. KGUW_ECOLI P24234 guanylate kinase (207 aa) fasta scores, opt: 424, E(): 6.6e-20, (35.9% identity in 184 aa overlap); contains PS00017 ATP/GTP-binding site motif A (P-loop), PS00856 Guanylate kinase signature"
  /codon_start=1
  /transl_table=11
  /product="gmk"
  /protein_id="CAB02172.1"
  /db_xref="GI:1542897"
  /db_xref="SWISS-PROT:P71659"
  /translation="MSVGECPDKTRKAGOPANVRVYVLSGPSAVGKSTVRCLEKRIPLNLFVSATTRAPRPGVBDVDYHFIDPTRFQOLIDQGLLENAETLHGGLHRSGLTAQPVRAAATGVPVLIENVLAGARAIKRTMPAVVFLAPSPWDLQARLLIGRTETA DVIQRLDPTARTEILAAGDDPKVAVVNRRLSEKCAELVSLVGTAGSP"
  4740..4763
  /gene="gmk"
  /note="PS00017 ATP/GTP-binding site motif A (P-loop)"
  4821..4874
  /gene="gmk"
  /note="PS00856 Guanylate kinase signature"
  5337..5341
  /note="possible RBS, AGGAG, for Rv1390"
```


FEATURES	
source	location/Qualifiers
misc_feature	1..29550 /organism="Mycobacterium tuberculosis"
	/strain="H37Rv" /db_xref="taxon:1773" <1..28776 /note="fragment designated v039. Does not represent a physical clone"
gene	complement(122..1147) /gene="gpda1"
CDS	complement(122..1147) /note="gpda1" /note="Rv0564c, (MTV039.02c), len: 341, gpda1, Probable dehydrogenase similar to Mycobacterium tuberculosis MTCY349.05 (334 aa), probable glycerol-3-phosphate dehydrogenase, and many other glycerol-3-phosphate dehydrogenases. Contains P500017ATP/GTP-binding site motif A (P-loop). FASTA scores:283018 MTCY349..5 (334 aa) opt: 740 z-score: 899.9 E(): 0; 40.4% identity in 322 aa overlap: and sp P46919 GPDA_BACSU GLYCEROL-3-PHOSPHATE DEHYDROGENASE (345 aa) opt: 731 z-score: 888.8 E(): 0; 37.3% identity in 332 aa overlap. Tbpase score is 0.891"
misc_feature	37..460 /codon_start=-1 /transl_table=-1 /product="gpda1" /protein_id="CAAI7435.1"
	/db_xref="GI:2909623" /db_xref="SPTREMBL:O53761"
gene	/translation="MAANKREPKVVIVGGSGWCTTVASICARRGPTLQWVSAVTAQD INDNHRNSRYLGDNDVYLDLIRATDFTPEANCAADVVMGVPSPHSGFVJVELSKELR PWPVYSLVGLGEGTMMRMSQIIEELVPGHAGIILAGPNIAREAVGVAALAEKLR DOHLATRLSMPFRFRFRVYTDVGVETAGALKVFAIANGVSLGAGCENTRALV IARALRMTKLYVAMGSGSEFPGLACGLIYCTQSRNRNRVQQLDAGRPIDDI IASMSQVAEKVAKAGVMEFRANFEGMLMPAREVDVAINHGSTEQAYRGLIAVPGH EVHSGCF"
misc_feature	complement(437..460) /gene="gpda1"
gene	/note="P500017 ATP/GTP-binding site motif A (P-loop)" complement(1208..2668) /gene="Rv0565c"
CDS	complement(1208..2668) /note="Rv0565c" /note="Rv0565c, (MTV039.03c), len: 486. Probable monooxygenase similar to other putative monooxygenases in Mycobacterium tuberculosis e.g. MTCY01A6.14 (489 aa), MYV013.4 (495 aa), MTCY31.20 (495 aa) and CYCLOHEXANONE MONOOXYGENASE P12015 CYMO_ACISP (542 aa) (EC 1.14.13..22) from Acinetobacter. FASTA scores: 283864 MTCY1A6.14 (489 aa) opt: 1571 z-score: 1903.6 E(): 0; 49.0% identity in 471 aa overlap: and sp P12015 CYMO_ACISP CYCLOHEXANONE MONOOXYGENASE (542 aa) opt: 354 z-score: 429.0 E(): 2.1e-16; 23.7% identity in 435 aa overlap. Tbpase score is 0.902"
misc_feature	435..460 /codon_start=-1 /transl_table=-1 /product="hypothetical protein Rv0565c"
	/protein_id="CAAI7436.1"
gene	/db_xref="GI:2909624" /db_xref="SPTREMBL:O53762"
CDS	/translation="MSVTPNAGCVDVIVVAGISGLGAAYRIERNPDLTYTILERRA RIGTWDLFRPGVRSDSITFLSPYEPREGIADGANIREYLDMAHKYCIDHH IENSVYRADADMSDTMTVTFEONGVNHYSREVFSGSYVNDGEGTPDGGIE KFGAVVHPQHPEDLDYTGKTIYVIGSCATVTLIPSLDRAKVTMLODRPYLLS ASKSTFAAYRKALPKPTSHLYIRYNALIEAVFWLSKTPFVYKMLRRTIKNL PEGIDIEHTHPKINPDORCLIPDADILNATISGRAEVVTHIDIEFDIALKSG GHLDDADIVATGLOLQAGAAISLDGEVIDPRDREVYKAHMLLEDVNPFWCGYTN ASWTLRADMTARATAKLILAHMAAGHTRAPAHGLGDEPDEKPSWDIQAGYVKAPVAL PKSGTRPNVVRONVYLADVIDYREDRIEAMVFGAA"
RBS	complement(2675..2678) /note="possible RBS, ggaag, for Rv0565c"
gene	complement(2746..3237) /gene="Rv0566c"
CDS	complement(2746..3237)
misc_feature	3237..422 /codon_start=-1 /transl_table=-1 /product="hypothetical protein Rv0566c"
	/protein_id="CAAI7437.1"
gene	/db_xref="GI:2909625" /db_xref="SPTREMBL:O53763"
CDS	/translation="MAQSDIVSKVDROGEVDNALNQAKELATREDGCTPTKIAWK GDEAVELSTEEVRKAAVVFKEKLRDISLKAPEGPOASGKYVYALCKQGI SEENAKITKLRDAGPKNKVTKIQCGEAVTSSKKRDLQAVIAMLKKADLVALQFV NYR"
RBS	complement(3248..3254) /note="possible RBS, aaggagg, for Rv0566c"
tRNA	3307..3387 /note="tyr, tRNA-Tyr : anticodon gta"
RBS	3507..3511 /note="possible RBS, ggaag, for Rv0567; possible RBS, aaggga, for Rv0569"
gene	3519..4538 /gene="Rv0567"
CDS	3519..4538 /note="Rv0567, (MTV039.05), len: 339. Probable methyltransferase similar to several e.g. TETRACENOMYCIN POLYKETIDE SYNTHESIS 8-O-METHYL TRANSFERASE sp P39896 TCMO_STRGA (339 aa). FASTA scores: sp P39896 TCMO_STRGA (339 aa) opt: 685 z-score: 836.6 E(): 0; 35.8% identity in 335 aa overlap:and sp P10950 H10M_BOVIN HYDROXYINDOLE O-METHYLTRANSFERASE(345 aa) opt: 509 z-score: 622.8 E(): 3.4e-27; 30.7% identity in332 aa overlap. Tbpase score is 0.910"
misc_feature	4538..4648 /codon_start=-1 /transl_table=-1 /product="hypothetical protein Rv0567"
	/protein_id="CAAI7438.1"
gene	/db_xref="GI:2909626" /db_xref="SPTREMBL:O53764"
CDS	/translation="WELSPRIMAVIGGYPSPKVLTLVAGLGLFTEIGDEPMAEAIA DRGLILRPAIDFDALVSLDLARDDGGSHVRYNTPERAHFLDEARPTVAGGLNLI WNERNYRPFMDLLEALTKYKAQSGSKVQTPFEALFADPRLEAFAAMDAASRRNI ELAKRPFERFRRLLCDVGCAGDLSRIVAAHPHLLQCVSFDPAAVEIARRKLTAG LGEHVQACGDFLADPLPADVITMGOILHDMNIDRQOVLARVAYALSREGAFIVE TLIDVARENTGLMSLNLIERGDAFDYSADFRMGCEAGFRFEVIFLAGSSA AVAYK"
gene	4648..6066 /gene="Rv0568"
CDS	4648..6066 /gene="Rv0568"
misc_feature	6066..6458 /note="Rv0568, (MTV039.06), len: 472. Unknown but similar to putative Mycobacterium tuberculosis cytochrome P-450 monooxygenase, MTCY63.32c (449 aa), and other cytochrome P-450 related enzymes e.g. P29980 CPYN_ANASP (459 aa). Contains cytochrome P450 cysteine heme-Iron ligand signature (P500086). FASTA scores: 296800 MTCY63.32 (449 aa) opt: 1080 z-score: 1266.4 E(): 0; 40.5% identity in 444 aa overlap: and sp P29980 CPYN_ANASP PROBABLE CYTOCHROME P450 (459 aa) opt: 525 z-score: 617.0 E(): 7.2e-27; 31.9% identity in 417 aa overlap. Tbpase score is 0.896"
	6458..6468 /codon_start=-1 /transl_table=-1 /product="hypothetical protein Rv0568"
gene	/protein_id="CAAI7439.1"
CDS	/db_xref="SPTREMBL:O53765"

misc_feature /db_xref="taxon:1773"
<1..66170
/note="fragment designated v017. Does not represent a physical clone"
RBS 178..181
/note="possible RBS, ggaag, for Rv1049"
gene 184..630
/gene="Rv1049"
184..630
/gene="Rv1049"
/note="Rv1049, (MTV017.02), len: 148. Probable repressor similar to many e.g. P74870 (149 aa). FASTA scores: sp|P74870|P74870NEGATIVE REGULATOR OF EMR LOCUS EMR (149 aa) opt:146 z-score: 200.3 E(): 0.0011: 31.6% identity in 95 aa overlap. Tbpase score is 0.892. Contains probable helix-turn-helix motif at aa 58-79 (score 1495, +4.28 SD)"
/codon_start=1
/transl_table=1
/product="hypothetical protein Rv1049"
/protein_id="CA17165.1"
/db_xref="GI:2896686"
/db_xref="SPTREMBL:O53397"
/translation="MGKAAFDKACACYTTRRAAQLQAVDRALRPSGLTNPQSTLA VISLSEGSAGIDLTMSLEAARIGVERTTLRNLEVMRRDGLVRVMAGADARCKRIELT AKGRAALOKAVPLMRGVQAEVTASVGDMPRVRRDIANLQGAEEACR"
RBS 668..672
/note="possible RBS, aggaag, for Rv1050"
gene 679..1584
/gene="Rv1050"
679..1584
/gene="Rv1050"
/note="Rv1050, (MTV017.03), len: 301. Probable oxidoreductase similar to many e.g. MTCY48.22C(341 aa). FASTA scores: sp|Q10783|Y04M_MCTU PUTATIVE OXIDOREDUCTASE C48.22C (341aa) opt:462 z-score: 533.9 E(): 3e-22: 33.6% identity in 265 aa overlap. Tbpase score is 0.896"
/codon_start=1
/transl_table=1
/product="hypothetical protein Rv1050"
/protein_id="CA17166.1"
/db_xref="GI:2896687"
/db_xref="SPTREMBL:O53398"
/translation="MARQRFQYVLLITGASSGIGETATKAFAREGAVALLAREGA LRRVAREIABAAGRAMVAPLDVSSSESVRAVADVGEFGRIDVFPNAGSIAVGPDA AETPLDTRREMLETDYLGTVAVREVLPIMKQORSGRIMNSSVYGRKAPRFQYSS AMHAIAGSDALROELKSGSIAVSVTHPALQTPLANVDPADMPPEPRLTPIPVH VAAVLDGVARRARVAVPEQPRLLMGDAFSPRYGDRVVRLLLESLTIFGRLTGYSRGS VYRHQPTESAKAQAAQPERGYSAR"
gene complement(1743..2498)
/gene="Rv1051c"
complement(1743..2498)
/gene="Rv1051c"
CDS /note="Rv1051c, (MTV017.04c), len: 251. Unknown but similar to protein encoded by Lactococcus lactis plasmid pNP40, LU036837.2 (298 aa). FASTA scores: U36837|LU036837_2 Lactococcus lactis plasmid pNP40, (298 aa) opt: 194 z-score: 245.5E(1): 5.5e-06: 30.3% identity in 155 aa overlap. Tbpase score is 0.912. Contains possible helix-turn-helix motif at aa197-218 (score 1097, +2.92 SD)"
/codon_start=1
/transl_table=1
/product="hypothetical protein Rv1051c"
/protein_id="CA17167.1"
/db_xref="GI:2896688"
/db_xref="SPTREMBL:O53399"
/translation="MRADVTAEHLTOVARDIAVIDIDGVAFNLDTSVQIEIRADY PGLRVAMASGYPWQGIANDVSTGEPIAPVTTIDRIIGEPITLTGAPETITAE KGVITLGRITSTWRDYVDIVLDKRGIDDDLEKRSARAAYRGATLEVAHLAG YGAVAAQKMATHEHRCQHCWRHMKPAHVGRRNMDDLDAKOVSEMIQVPGVTLRMRHS DIGPASTLGRVYVRDEVSRRWISKRESATRR"
RBS 3507..3513
/note="possible RBS, aaggtgg, for Rv1052"

gene 3521..3910
/gene="Rv1052"
CDS 3521..3910
/gene="Rv1052"
/note="Rv1052, (MTV017.05), len: 129. Unknown"
/codon_start=1
/transl_table=1
/product="hypothetical protein Rv1052"
/protein_id="CA17168.1"
/db_xref="GI:2896689"
/db_xref="SPTREMBL:O53400"
/translation="MDCCEGRVARHKGLSQVGTGCGPRMSQAVSCRCASVREAAVTA VOMPLTGYCGTPLPHDELAAIPEVEVDLKPITRADYVDLEQLODOVFDLMPFA VEGSLSDLELSDHFVRDLHARMGVP"
gene complement(3809..4084)
/gene="Rv1053c"
complement(3809..4084)
/gene="Rv1053c"
CDS /note="Rv1053c, (MTV017.06c), len: 91. Unknown"
/codon_start=1
/transl_table=1
/product="hypothetical protein Rv1053c"
/protein_id="CA17169.1"
/db_xref="GI:2896690"
/db_xref="SPTREMBL:O53401"
/translation="MDSHKVCMNNNTOLPTGPIIGVPAVRDGVRYAIIQDGLRCN TDVEFTSSPEPGPVLYTKTKTRVRIADDMVTEKLKRQAFNSRRIQ"
gene 4726..4932
/gene="Rv1054"
4726..4932
/gene="Rv1054"
/note="Rv1054, (MTV017.07), len: 68. Unknown but similar to hypothetical Mycobacterium tuberculosis protein MTCY3G12_25 (151 aa shows similarity to integrases) and to Mycobacterium paratuberculosis integrase MSGINT_1 (191 aa). This orf continues in another frame as MTV017.08 but no error can be found to account for frameshift. FASTA scores: 279702|MTCY3G12.25 (151 aa) opt: 273 z-score: 364.1 E(): 8.8e-13: 64.7% identity in 66 aa overlap; and L39071|MSGINT_1 (191aa)opt: 105 z-score: 148.5 E(): 0.9; 31.8% identity in 85 aa overlap."
/codon_start=1
/transl_table=1
/product="hypothetical protein Rv1054"
/protein_id="CA17170.1"
/db_xref="GI:2896691"
/db_xref="SPTREMBL:O53402"
/translation="MTGKGIVESTTKTKRDHVPVPEPVMRRLHAELPTDPNALVPPG RKGGFLPLGEYRMAFDNAGQVQITE"
gene 4935..5171
/gene="Rv1055"
4935..5171
/gene="Rv1055"
CDS /note="Rv1055, (MTV017.08), len: 78. Partial orf, first 49aa similar to hypothetical Mycobacterium tuberculosis protein MTCY3G12_25 (151 aa shows similarity to integrases) and to Mycobacterium paratuberculosis integrase L39071|MSGINT_1 (191 aa)and to many other integrases or transposases. FASTAscores:279702|MTCY3G12.25 (151 aa) opt: 291 z-score:428.6 E():2.2e-16: 74.3% identity in 70 aa overlap; andp|L39071|MSGINT_1 (191 aa) opt: 146 z-score: 221.0 E(): 8.3e-05: 52.1% identity in 48 aa overlap"
/codon_start=1
/transl_table=1
/product="hypothetical protein Rv1055"
/protein_id="CA17171.1"
/db_xref="GI:2896692"
/db_xref="SPTREMBL:O53403"
/translation="MVPHGIGHTTASLAISGANVKKVQVORLLGHAAMTLDIRGHILL NDDLAIVPMRCASSFTLRVHCGRRRRRVGLRA"
TRNA 5194..5267

aa). Contains inositol monophosphatase family signature 1 (P500629), significance uncertain. FASTA best: CY50_MUCLE P46726 cy5q protein homolog (289 aa) opt: 1374 z-score: 1597.06(E): 0; (77.3% identity in 264 aa overlap)"

/codon_start=1
/transl_table=1
/product="cy5q"
/protein_id="CAB08639.1"
/db_xref="GI:2104347"
/db_xref="SPTREMBL:006244"
/translation="MVSAPADLTDDLTDAELADLAADAGKILLQVRAEIGFDQPMF LGEAGDQANSLRLRLQAEPPGADVAESEAHDLARIKSPRWIIPDLGTRFSTP GRDMMVAHIALMRSSNGOPEITDAVALPARGNVVYRTDVTSGAAPAGPGRLRA VSAIRPAAVLRHRIQTLAIOPVSTISGAKAMAYIDGVAYLIHAGCOMEDSAAAPAG VMLAAGHASTKDSPLRYNDLPDLPDLMLCRAEVAPILGALADAMR"
complement(552. .593)
/gene="cy50"
/note="P500629 Inositol monophosphatase family signature 1"
945. 1175
/gene="Rv2132"
945. 1175
/gene="Rv2132"
/note="Rv2132, (MTCY270.36c), len: 76. Function unknown but some similarity to Mycobacterium tuberculosis protein A1021924(MTV020.4 (84 aa). FASTA score: opt: 142 z-score: 210.0 E(): 0.00036; 41.8% identity in 55 aa overlap)"
/codon_start=1
/transl_table=1
/product="hypothetical protein Rv2132"
/protein_id="CAB08638.1"
/db_xref="GI:2104346"
/db_xref="SPTREMBL:006243"
/translation="MRTTVSLADVAAPVLRKERSIGLSEAVNELIRAGLTKRQVA NRFOQYDMGEGIDYSNIGDAIETLDPGASG"
complement(1385. .2173)
/gene="Rv2133c"
complement(1385. .2173)
/gene="Rv2133c"
/note="Rv2133c, (MTCY270.35), len: 262. Function: unknown but equivalent to hypothetical M. leprae protein, Q49774. FASTA best: Q49774 B2126_C2.150. (262 aa) opt: 1447 z-score: 1647.5 E(): 0; (79.0% identity in 262 aa overlap)"
/codon_start=1
/transl_table=1
/product="hypothetical protein Rv2133c"
/protein_id="CAB08660.1"
/db_xref="GI:2104345"
/db_xref="SPTREMBL:006242"
/translation="MLADGELTVLGRIRASNAITFLCESTLGLRLHLCVYKPVSGERP LMDPDTLAGERELASAVLSTQLGMNVPHITIRDPAGIGMLHMQPQDAVDSP LPPDLVDFEAPHRPGYLVLRAYDAGVEVLMHADDIRLRMAVFFDLINNAIR KGHILGIDQVYGDHGLCEHVENKLRITVLMGMAKCPIDDOIILQAVAGLADALGSP LAELAGRIAAEIGALRRRQSLDDQVPMGPGNHRPIMPAP"
complement(2184. .2771)
/gene="Rv2134c"
complement(2184. .2771)
/gene="Rv2134c"
/note="Rv2134c, (MTCY270.34), len: 195. Function: unknown but equivalent to hypothetical M. leprae protein, Q49789. FASTA best: Q49789 B2126_C3.228. opt: 1192 z-score: 1493.1 E(): 0 (91.1% identity in 192 aa overlap)"
/codon_start=1
/transl_table=1
/product="hypothetical protein Rv2134c"
/protein_id="CAB08659.1"
/db_xref="GI:2104344"
/db_xref="SPTREMBL:006241"
/translation="MARAHVFRTPDRFVAGTYGQPCNRRFFYLQAVHDSRVSVLEK QQVAVLAIRGALLFEVNRRRGTPVPEPEITDLSPLIMVDAEFVNGTGLMDSB AOSVVELAVTADEFAVSVDLDEGSPAPVAVRIPESAKRQFATRSYRIVISAGRRP CPLDEPLDEGHICARTNGYRVDVLLGSGDDPAG"
complement(2776. .2780)

/note="possible RBS, GGAGC, for Rv2134c"
complement(2835. .3545)
/gene="Rv2135c"
complement(2835. .3545)
/gene="Rv2135c"
/note="Rv2135c, (MTCY270.33), len: 236. Function: unknown but equivalent to hypothetical M. leprae protein, Q49773. FASTA best: Q49773 B2126_C1.148 opt: 1183 z-score: 1390.9 E(): 0; (74.8% identity in 250 aa overlap), also similar in C-terminus to PwG2_ECOLI P36942 probable phosphoglycerate mutase 2 (215 aa) fasta scores: opt: 212 z-score: 257.4 E(): 2.5e-07 27.9% identity in 190 aa overlap"
/codon_start=1
/transl_table=1
/product="hypothetical protein Rv2135c"
/protein_id="CAB08658.1"
/db_xref="GI:2104343"
/db_xref="SPTREMBL:006240"
/translation="MTYILRHARSTNTAGVLAGRSVDLDEKGRQATGLIDRIGD LPIRAVASSPMLRCQRTVEPLAEALCLEPILIDRFSEVDYGEWTKRIGDLVEPLMR VQAHPSAAVFPGGESLAQVOTRAVAARREHRLADHGVLMIACTHGDVTKAYT ADAFGMLDSFORITADPGSVSVRYTLRPFLVNHMGARLAPALQAAASAQGASP EPNAAVPPGDAVIGGSYD"
complement(3542. .4372)
/gene="Rv2136c"
complement(3542. .4372)
/gene="Rv2136c"
/note="Rv2136c, (MTCY270.32), len: 276. Function: unknown possible membrane protein very similar to hypothetical M. leprae protein Q49783. FASTA best: Q49783 B2126_C2.190. opt:1023 z-score: 1259.2 E(): 0; (82.4% identity in 187 aa overlap) similar to BACA_ECOLI P31054 bacitracin resistance protein (273 aa) opt: 477 z-score: 590.1 E(): 7e-26, (35.6% identity in 267 aa overlap)"
/codon_start=1
/transl_table=1
/product="hypothetical protein Rv2136c"
/protein_id="CAB08657.1"
/db_xref="GI:2104342"
/db_xref="SPTREMBL:006239"
/translation="MSMWOYIVLAAAGLHEFLPVSSSGHIAIVRIFSGDAGASPT AVSOLGFAAVVYIFARDIVRIISAMHGVVAHRQSDRLGMYVITGIPICITGL FPKDIDRSQVRNLMVYVLTALVPSGVALIALEYGVRGRHIERLRRAVVGIAQTILA LVPGVSRSSGSTITISAGFLGIDRELAARFLAIIPAFASGLSFLPAFHPVEGMSA TGPDLVATLIAFVLGILTAVALMLRFLVRHMYVGVRYLVGTGMVLATGTVAAVT"
complement(4436. .4849)
/gene="Rv2137c"
complement(4436. .4849)
/gene="Rv2137c"
/note="Rv2137c, (MTCY270.31), len: 137. Function: unknown C-terminus is very similar to hypothetical M. leprae protein B2126_C2.188 (150 aa). FASTA best: Q49782 B2126_C2.188. (150 aa) opt: 469 z-score: 623.7 E(): 9.6e-28; (77.2% identity in 101 aa overlap)"
/codon_start=1
/transl_table=1
/product="hypothetical protein Rv2137c"
/protein_id="CAB08656.1"
/db_xref="GI:2104341"
/db_xref="SPTREMBL:006238"
/translation="MRNMKSTSHSESGKLSTSSCRPREMYLQRSYLGMTVYADRHL ADKREFAVEDISTGIFPASGYGVGDGRSFSFHIHRSILVEIYRPAVAGPVAQEDV

Query Match 29.3% Score 340.6; DB 2; Length 37586;
Best Local Similarity 61.7%; Pred. No. 4.8e-15;
Matches 737; Conservative 0; Mismatches 404; Indels 54; Gaps 10;

QY 4 togatgccagtcgccgcgcgcagccagccgctcacgatacaacccgcggtgcg 63;
||| ||| ||||| ||| || ||||| ||||| ||| || |||||
DB 30955 TCGCCGCTTCGCCGCGCTTGCGCGCGCGCGCTACGCAGATCAGCGTGGCGG 31014

RBS

[illegible][illegible]

[illegible][illegible]


```
QY 65 -----ALEAPNPVPAPGPG-NSA--PAPMP-- 90
DB 350 CCRVSGEVLGALNPSRPSPLTTTPALPAPLPPLPPLINAVPIPLPPLV 409
QY 91 -----TPPLPVPGSGABPVPVPMMPAKRMALPPA----- 126
DB 410 ALAPPLPAPLPISGVPAPAPLIPGK--PWTTPPLAPAPPEPKTVLPLGSCPSSE 467
QY 127 ---PPAPSPPTSLAVPVVPVPPVPLP--VKMPSPVPVPFPAPPEET-----P 172
DB 468 KPNBPAPPEPEPKSSPALPPAPAPMSAVRVPPSPDIPAPAAERASMPALPPAP 527
QY 173 NPAP-----PPAPPLNSPVPVPVPPV-----PPVPLTLNPPVPPAPPA- 212
DB 528 SPPTATRLCPPLPSPAPANSPPAPAPAPPTPPKLSANPCCPVPAPAPKRPAPAPAP 587
QY 213 -----ANTNSPLRPPAPAPAPPLK---GCPPAPMPAPANS---PAPSPSP 255
DB 588 ELAPAPDPPTPPVANSPPAPAPAPAPALPFPVNPAPPTPAAPKSRPALPAPAPAP 647
QY 256 PVPVFPPTPPGPPAPPEPNS--PPAPAPAPAPLPGSPAPAPAPPLNSPAPAPGPPAW 313
DB 648 PVRAATTPAPAPAPAPAPAPAPAPAPAPPLP--ATPPAPAPAPLPMSPAPAPLPPA- 704
QY 314 PGADPPAPPLPYSPPAPACVPAGAPLAPLISGRP-----SNSWGVFTMLSRPSNGA 369
DB 705 --ADPPAPPLTLNPPSPPLAPVPGALPLPINGRVFAKKNLLIG-----SSSGD 755
QY 370 AAASALA 377
DB 756 TAAASAAA 763
RESULT 2
Q9SBM1 PRELIMINARY; PRT; 409 AA.
AC Q9SBM1;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2000 (TREMBlrel. 14, Last annotation update)
DE HYROXYPROLINE-RICH GLYCOPROTEIN DZ-HRGP PRECURSOR.
GN HRGP GENE.
OS Volvox carteri f. nagariensis.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Volvocaceae; Volvox.
OX NCBI_TaxID=3068;
RN [1]
SEQUENCE FROM N.A.
STRAIN-HK10;
Ender F., Hallmann A., Amon P., Sumper M.;
"Response to the sexual pheromone and wounding in the green alga
Volvox: induction of an extracellular glycoprotein consisting almost
exclusively of hydroxyproline.";
RL J. Biol. Chem. 274:35023-35028(1999).
DR EMBL; AJ242540; CAB62280.1; -.
DR INTERPRO; IPR002965; -.
DR INTERPRO; IPR002966; -.
DR INTERPRO; IPR003072; -.
DR PRINTS; PR00239; RHODOPSNTAIL.
DR PRINTS; PR01217; PRICHEXTENSIN.
DR PRINTS; PR01218; PSTLEXTENSIN.
DR PRINTS; PR01286; NORNUCRECEPTR.
KW Signal.
FT SIGNAL.
FT CHAIN
SEQUENCE 409 AA; 41547 MW; CD0749C6AF02BD74 CRC64;
```

```
QY 6 PPAPAPSPINPVPPVPLPAPARTLSPPVAPSPISLAAPLPDPMPALWISA 65
DB 66 PPPPPPPPP--PPQPLPSPSPSPPPPPPPPPPPPPPP--PPPPPPSPBPPPP 115
QY 66 LEADNPVPAPGNSAPAPAPMPPTPLPVPVPGSAPAPVAPVPMAPAKRMALPP 125
DB 116 --PPPPPPSPPPPPPPPPPPPPPPPPPPPPPP--PSPPSPSPSP--PSPPSPSP 170
QY 126 APP--APPSPPTSLAVPVVPVPPVPLP--VKMPSPVPVPFPAPPEET-----P 179
DB 171 SPSPSPSPSP-----PSPPPPPPSPSPSPSPSPSP--PPPPPPSPSPSPSPSP 223
QY 180 PPLNSP--PPPVPPVPVPLTLNPPVPPAPAPANSNSPLRPPAP--PAPPLKGP-PA 236
DB 224 PP--SRPPPPRRPPPPPPPPPPPP--PPAP-----PPRRPPSPPLPPLPPSP 272
QY 237 PPMAPAPNSPAPSPSP--PVPVFPPTPGP-----PAPPEPNSPPAPPA---PPA 284
DB 273 PPLPPSPSP--PPSPPPSPPLPSPPPAPKPSPPNPFPPPPPPPPPPPPSPMPPR 330
QY 285 APPLGSPAPAPAPPLNSPAPAPGPPAWGAPPPA--PPLPYSPPAPACVPAPLA 343
DB 331 PPPSPPPPP--PPSPVPPPPPPSPSPKPPPPSPPPPPPPPPPPPPPP--PTA 385
QY 344 PLPISGRPSNSW 355
DB 386 PGVVGFPDPDEM 397
RESULT 3
Q9TOK5 PRELIMINARY; PRT; 760 AA.
AC Q9TOK5;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE EXTENSIN-LIKE PROTEIN.
GN T9E8.80 OR AT4G13340.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
SEQUENCE FROM N.A.
RP Bevan M., Pohl T., Weizengger T., Bancroft I., Mewes H.W.,
RA Mayer K.F.X., Schueller C.;
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
SEQUENCE FROM N.A.
RP EU Arabidopsis sequencing project;
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
SEQUENCE FROM N.A.
RP Pohl T., Weizengger T., Mewes H.W., Lemcke K., Mayer K.F.X.;
RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
SEQUENCE OF 348-747 FROM N.A.
RA Robben J., Grymonprez B., Volckaert G., Mewes H.W., Lemcke K.,
RA Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL049608; CAB40769.1; -.
DR EMBL; AL161536; CAB78376.1; -.
DR INTERPRO; IPR002965; -.
DR INTERPRO; IPR002966; -.
DR PRINTS; PR01217; PRICHEXTENSIN.
DR PRINTS; PR01218; PSTLEXTENSIN.
SEQUENCE 760 AA; 82245 MW; 20CD1C8E6CECFC CRC64;
```

Query Match 37.5%; Score 811; DB 10; Length 409;
Best Local Similarity 48.7%; Pred. No. 1e-52;
Matches 181; Conservative 15; Mismatches 114; Indels 62; Gaps 22;

Query Match 31.0%; Score 687; DB 10; Length 760;
 Best Local Similarity 42.1%; Pred. No. 3.4e-42;
 Matches 164; Conservative 30; Mismatches 112; Indels 84; Gaps 24;

QY 5 VPPAPAPSPDINPVVPPPLPA-----APRTLSPPVPPAPSPDLSLAAPLPDPMP 60
 Db 392 VSPRPV-VIPLPPSLSPSPRPAPFISTPPTLSPPSPSPPVV--YSP--PPPPPP 446
 QY 61 AIMSALAPNPVPPAPPGNSAPAPMPTPLPV--PPSGAPRPVAPVMPAPK 118
 Db 447 PVYS-----PPPPPPPPPPVYSSPPPPPPPPPVYSP-----PPSPPPPPPV 494
 QY 119 RMLPAPAPAPSPSPMLAVPVV--PPVPLPVKMP-----SPPPPPPPAPPE 170
 Db 495 YSP--PPPPPPPPPVV--YSPPPVYSSPPSPAPFVYCTRPPPPHSPPPPPQF 550
 QY 171 TPMPAP-----PAPPLENSPPPPPPVPPVPL--TLNPVPPAPPAANTSPLRP 223
 Db 551 SPPEPEVYSSPPPPHSPPPHSPPPHSPPPHSPPPHSPPPHSPPPHSPPPHSP 609
 QY 224 APPAPPLKPGPPAP-----PMPAP-----NSPAP-----SPSPVYVETPPGPP-- 267
 Db 610 PPEPPCIEPPPPPPCIEYSPPPPPVYHSSPPPPVYSSPPPPVYSSPPPPPPV 669
 QY 268 -APPEP-----NSPP-----APPAPAPLPSPSPAP--PAPPLPNSPAPAPGPPAP 313
 Db 670 SSPPPPVYHSSPPSPVYHSSPPSPAPCESPPAPVYHSSPPPPVYHSSPP 728
 QY 314 PGAPDPAP-----PLP-----YSSPPAP 333
 Db 729 HQSPPPSPPEGLPPLVIGVASYASPPPP 758

RESULT 4
 ID Q41805 PRELIMINARY; PRT; 1188 AA.
 AC Q41805;
 DT 01-NOV-1996 (TEMBUREL. 01, Created)
 DT 01-NOV-1996 (TEMBUREL. 01, Last sequence update)
 DT 01-JUN-2000 (TEMBUREL. 14, Last annotation update)
 DE EXTENSIN-LIKE PROTEIN PRECURSOR.
 OS Zea mays (Maize).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; Liliopsida; Poales; Poaceae; Zea.
 OX NCBI_TaxID=4577;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA RUBINSTEIN A.L., Broadwater A.H., Lowrey K., Bedinger P.A.;
 RL Submitted (JUN-1994) to the EMBL/GenBank/DBJ databases.
 DR EMBL: Z34465; CAAG4230.1;
 DR MENDEL: 14346; Zeama; 2368; 14346.
 DR INTERPRO: IPR001611;
 DR PIRAM: PF00560; LRR; 3.
 KW Signal.
 FT SIGNAL 1
 SQ SEQUENCE 1188 AA; 120980 MM; 2C77C7F8D7130149 CRC64;

Query Match 30.3%; Score 670.5; DB 10; Length 1188;
 Best Local Similarity 32.1%; Pred. No. 7.2e-41;
 Matches 190; Conservative 37; Mismatches 149; Indels 215; Gaps 28;

QY 2 SMVP-----PAPAPSPDINPVVPPPLPAAPRTLSPPVPPAPSPDLSLAAPLPDPMP 58
 Db 524 SPAPAGSP 583
 QY 59 PPAIWSALEAPNPV--PPAPPGNSAPAPM-----PPTPLPVVPGSGAPRPVAPVMP 113
 Db 584 PPL--V--VASPPPVKSPPPAPVAVASPPPVKSPPTPVASPPAPVAVASPPPMKSP 640
 QY 114 PPAKRMPPALP-----PAPAP-----PSPPTSLAVPV-----PV-- 146

Db 641 PPTVSSPPPEKSPPPPPAKSTPPPEEYPTPTSTVSKSSPPPKSLPPLILSPPEQ 700
 QY 147 -PVPLPVKMPSPSPVPFPAPAE-----PETPN--PDA-----PPA----- 179
 Db 701 KPPTSTPKSPKSPSPKSPKSPKSPKSPKSPKSPKSPKSPKSPKSPKSPKSP 760
 QY 180 -PLENSPPPPPVVPPVPPPLTNP-----VPPA-----PPA 212
 Db 761 PPSVSKSSPPAPLSSPPAPQVSKSPPVYSSPPPPAKSSPLAPVSPPOYEKTSPP 820
 QY 213 ANTSNPLR-----PPAP-PAPPLKPGPAP----- 238
 Db 821 APSSPPLAKSSPPPHVSSPPPVVSSPPPPAPVSSPPLTPKAPSAHVAVSSPPEVVK 880
 QY 239 -MPAPNSPAPSPSPS--PPVPVFPPTP-----GPPA----- 268
 Db 881 STPPAPTTVSPSEPKSPSPPTFVSLPPLVSKSPPPAMVSSPMTPKSSPPVYSSP 940
 QY 269 -----PPEPNSPPA-----PPAPPAAPLPGP 290
 Db 941 PPTVSKSPPPAPVSSPPATRKSSPPAPVNLPPPEVKSPPPTPVSSPPAPKSSPPAP 1000
 QY 291 -SPPAP-----PAPPLP-NSPAP-----GPPAMGAPDPA--PPLPYSSPPAPACP 336
 Db 1001 MSSPPPEVKSPPPPAPVSSPPPPVKSPPPPAPVSSPPPVKSPPPAPVSSPPPVKSP 1060
 QY 337 VPGAPLA--PLPISGRPSNMGVETMLSRPSNGAASALAVAPAPAVK 385
 Db 1061 PPAPISSPPPPVKSPPPPAPVS-----SPPPVKSPPPPAPVSSPPPIK 1106

RESULT 5
 ID Q41707 PRELIMINARY; PRT; 489 AA.
 AC Q41707; Q43689;
 DT 01-NOV-1996 (TEMBUREL. 01, Created)
 DT 01-NOV-1996 (TEMBUREL. 01, Last sequence update)
 DT 01-JUN-2000 (TEMBUREL. 14, Last annotation update)
 DE EXTENSIN CLASS I PROTEIN PRECURSOR (EXTENSIN-LIKE PROTEIN).
 GN EXT26G.
 OS Vigna unguiculata (Cowpea).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
 OC Fabales; Fabaceae; Papilionoideae; Vigna.
 OX NCBI_TaxID=3917;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97155574; PubMed=9002273;
 RA Arsenijevic-Maksimovic I., Broughton W.J., Krause A.;
 RT "Rhizobia modulate root-hair-specific expression of extensin genes";
 RL Mol. Plant Microbe Interact. 10:95-101(1997).
 DR EMBL: X91836; CAA62943.1;
 DR EMBL: X86030; CAA60022.1;
 DR MENDEL: 16557; Vignu; 2930; 16557.
 KW Signal.
 FT SIGNAL 1
 SQ SEQUENCE 489 AA; 53896 MM; E927EF520CE258B3 CRC64;

Query Match 29.7%; Score 659; DB 10; Length 489;
 Best Local Similarity 40.4%; Pred. No. 2.5e-40;
 Matches 166; Conservative 27; Mismatches 144; Indels 74; Gaps 27;

QY 4 PVPPAP-----PAPAPSPDINPVV-----PVPLPAAPRTLS--SPVP--PAPPSPLSLA 49
 Db 86 PSPPPPVYSSPPPPSP--SPPPVYKSPSPSPSPSPSPSPSPSPSPSPSPSPSP 144
 QY 50 PPLPDPMPAPPAIWSALEAPNPVP-----APPGNSAPAP--MPPTPLPVVPG- 100
 Db 145 PPPSPSPSPPPVYKSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 204
 QY 101 --SGAPRPVAVP-----MPAPKRMPPALP-----PAPAPSPPTSLAVPVVPPV-P 147

Query Match 29.1%; Score 645.5; DB 10; Length 1315;
 Best Local Similarity 35.2%; Pred. No. 4.8e-39;
 Matches 172; Conservative 42; Mismatches 145; Indels 129; Gaps 28;

QY 2 SMFVPPAPAP-----PSPIMPPV-----PVPLPA- 28
 DB 428 SVGCKAPAPAKMPTPTPPDVSSPEPLPEPSPVAPAPAKMPTLRSPADEYIPTPVPAK 487
 QY 29 APRTLSPP-----VPAAPSPISLAAPLPDPPMPALWSAL-----EAPMPVP-----PAP 77
 DB 488 SPGCTSPSPASRGAPLQAOAPPAASSPPATPVKSSPPAAVLLPPAPATSPSPAPVASPP 547
 QY 78 PGNASAPAP-----PMPTPLPPVPPGSGAPAP--VPAVPPM-----PPAPKMPALPPA 126
 DB 548 EAPVSSQPOVKSPPPAPVAPASPPPMKSSPPPARVASPPPLKSSPPAP--VASPPQ 604
 QY 127 PPAPSPPTSMALAVPVVPVPPVPLPVKMPSPVPPEPAPAP-ETPNPPA--PPAP-PL 182
 DB 605 PLKSPPPVLMSTPSVKSPP-PPVPAVASPP-PPVKSPPPLAVSSPSPPVKLPLPAPG 662
 QY 183 ENSPPPPPPVPPVPPPLTINPP-----VPAAPPAANTSNSLRPPAPAPPLKGPAPAP- 238
 DB 663 KSTPPPEEEKPTPTPVKSSPPEKSLP--PPLTTSPPQEKPTPTPTSKPPSPVE 720
 QY 239 --MPPAPNSP-----AAPSP--SPVVPPTPPGPPAPAP--PNSSPPAP--APPA 285
 DB 721 TLPPPKSSPEEPVSSPPAPKSSSPAPVSSPPPLKSSPPVPESSPPPTKSSPLPA 780
 QY 286 PLRGP-----SPAPAPAPPLNPSAPAP-----GPPAMGAPPP--A 321
 DB 781 PVSSPPQVEXTSPPAVSSPPPTPKSSPLAVSSPPQVEXTSPPAVSSPPPTKSS 840
 QY 322 PPL-PISSPP-----APPACVPGAPLAPLPISGRPSNSWGVFTMISRNSGAAAAASA 375
 DB 841 PPLAPVSSPPQVEXTSPPAVSSPPLEPKP-----SSPSSVSSPPTT 884
 QY 376 LAYAPAPA 383
 DB 885 VKSSPPPA 892

RESULT 8
 ID 065553 PRELIMINARY; PRT; 3247 AA.
 AC 065553;
 DT 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
 DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)
 UL36.
 OS Bovine herpesvirus 1.
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Alphaherpesvirinae; Varicellovirus.
 OC NCBI_TaxID=10320;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=COOPER;
 RA Schwyzer M., Vicek C., Lowery D.E., Bello L.J., Meyer G., Misra V.,
 RA Labolisiere S., Misra V., Vicek C., Paces V.,
 RL Vet. Microbiol. 0:0-0(1996).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=COOPER;
 RA Schwyzer M.;
 RL Submitted (Aug-1996) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE OF 1-179 FROM N.A.

RC STRAIN=JURA;
 RA MEDLINE=97164286; PubMed=9010999;
 RA Schwyzer M., Stryger D., Vogt B., Lowery D.E., Simard C.,
 RA Labolisiere S., Misra V., Vicek C., Paces V.,
 RT "Gene contents in a 31-kb segment at the left genome end of bovine
 RT herpesvirus-1.";
 RL Vet. Microbiol. 53:67-77(1996).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN=JURA;
 RA Schwyzer M., Vicek C., Lowery D.E., Bello L.J., Meyer G., Misra V.,
 RL Submitted (Nov-1997) to the EMBL/GenBank/DBJ databases.
 RN [6]
 RP SEQUENCE FROM N.A.
 RC STRAIN=JURA;
 RA Schwyzer M.;
 RL Submitted (Sep-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; 278205; CAB01605.1; -;
 DR EMBL; AJ004801; CAA06097.1; -;
 SQ SEQUENCE 3247 AA; 332190 MW; 3AEAA72F8F001F6A CRC64;

Query Match 29.0%; Score 642.5; DB 12; Length 3247;
 Best Local Similarity 34.2%; Pred. No. 1.6e-38;
 Matches 195; Conservative 20; Mismatches 145; Indels 211; Gaps 31;

QY 2 SMFV-----PP-----APAPSPPTINPPVPPVPLPAAPRTLSPPVPPAP--SPIS 46
 DB 2593 SMFAALPWRRRPPASLVSAPVPPGRLPPAPPLP--PAP--LPPAPPLPPAPPLP 2649
 QY 47 LAAPPLPPD--PMPAPLWSLEAPNPPVPPAPPGNAPAP--PMP--PTPLP--PVPPSGS 102
 DB 2650 PPAPPLPPAPPLPPLPPASAPVAPVAPPLPALTPALTPAPTAPPLPPLPAPITVL 2709
 QY 103 APPVPAVPPMP--PAPKMPALPPAPAPPSPP----- 134
 DB 2710 VPAPVPAPIPAAPPTAPAPPTAPPLPPAPDGAAGALSATRRPTRRAGAKKSUPAAQ 2769
 QY 135 -----TSMALVPV-----PVP--PVPPPVKMP 156
 DB 2770 PROTLRSRSPASVAPAGSELVPPSGALGSPSPFVSPRPSLEPPGILPPSRVOAP 2829
 QY 157 ---PSPVPPPPAPPETPMPAPAPPLNPSPPPPVPPV--PVPLTLNPPVPPAPPA 212
 DB 2830 VDAAPPPAPERPAPPPAPPPAPPPAPPPAPPPAPPPAPPPAPPPAPPPAPPPAP 2889
 QY 213 ANTSNSPLRPPAP-----PAPPLKGPAPAP----- 238
 DB 2890 AVSGSKT--PAPSNALVGVAVALLRGSPGKPCAPAPRTPGGANIARKNTRAVDPTV 2947
 QY 239 -----MPAPNSPAPPSPPSPVPPVPPPGPPAPPEP 272
 DB 2948 IGSGALSGLGSAEDADFGASGMVYAPPGAPLAPAPASALPAPLEAPAAASLPPAP 3007
 QY 273 NSSPPAP--APPAALPUGSPAPAP--APPLNPSAPAP--GPPAMPGADPPAPPLVSSP 329
 DB 3008 IAPPIAPPIAPPIAP-----PIAPPIAPPIA--PIAPPIAPPPAP--APPSAAPTAGSSA 3060
 QY 330 P--APPACVP--GAPL--APLP-----ISGR-----PSNSWGV 358
 DB 3061 PPLAPPTAPPLAPLAPLAPVPLVGAARGRVIGHMMAAPKRGGLAARPPMRMSMA-- 3118
 QY 359 FTMLSRPSNG-----AAAAASALAVAPAP 382
 DB 3119 -----SRCLHTODILASAEARICELPAP 3142

RESULT 9
 ID 09NID1 PRELIMINARY; PRT; 998 AA.
 AC 09NID1;
 DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)

DT 01-OCN-2000 (TREMBlrel. 15, Last annotation update)
 DE CARBOXYL-ESTER LIPASE.
 GN CEL.
 OS Gorilla gorilla (gorilla).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Gorilla.
 OX NCBI_TaxID=9593;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20018178; PubMed=10548728;
 RA Madeski K., Lidberg U., Bjursell G., Nilsson J.;
 RT "Characterization of the gorilla carboxyl ester lipase locus, and the
 RT appearance of the carboxyl ester lipase pseudogene during primate
 RT evolution.";
 RL Gene 239:273-282(1999).
 DR EMBL: AF206618; AAF71700.1; -
 SO SEQUENCE 998 AA; 101025 MW; 3C81AD44504E8ABA CRC64;

Query Match 28.8%; Score 638.5; DB 6; Length 998;
 Best Local Similarity 41.3%; Pred. No. 1.2e-38;

Matches 181; Conservative 29; Mismatches 115; Indels 113; Gaps 35;

QY 4 PVP-----APPAPSPIN--PVP-----VPLPAAPRTLSPPV-----APPSPIS 46
 ||||| ||||| : ||||| : ||||| ||||| :
 DB 562 PVPPTGASEAPVPTGDSGAPVPTGDSGAPVPTGDSGAPVPTGDSGAPVPTGDSGAPVPTGDS 621
 QY 47 --LAAPLPD-----PMPALMSALEAPNPVPP-----APCP--NSAPAPMPPT- 91
 ||||| ||||| : ||||| ||||| : ||||| ||||| :
 DB 622 GDSGAPVPTGDSGAPVPTGDSGAPVPTGDSGAPVPTGDSGAPVPTGDSGAPVPTGDS 677
 QY 92 -PPLPVPVP--GSGAPRPV-----AVPMPAPKRMPLPAP-----APSPPTSL 138
 ||||| ||||| : ||||| ||||| : ||||| ||||| :
 DB 678 DSGAPVPTGDSGAP--VPTGDSGAPVPTGDSGAPVPTGDSGAPVPTGDSGAPVPTGDS 734
 QY 139 -AVPVP-----VPPVPL--PVKMPSP-----VPPPEAPETNPAPAPLENSP 186
 ||||| ||||| : ||||| ||||| : ||||| ||||| :
 DB 735 GAPVPTGDSGAPVPTGDSGAPVPTGDSGAPVPTGDSGAPVPTGDSGAPVPTGDS 789
 QY 187 PPPVPP-----VPPVPL--TLNPVPP-----APPAANTSNLRLPAP-----APP 229
 ||||| ||||| : ||||| ||||| : ||||| ||||| :
 DB 790 GAPVPTGDSGAPVPTGDSGAPVPTGDSGAPVPTGDSGAPVPTGDSGAPVPTGDSGAP 848
 QY 230 LKP--GPPAPMPAPNSPAPSP-----SPVPVPTTGPAPAPSPNS--SPAP- 280
 ||||| ||||| : ||||| ||||| : ||||| ||||| :
 DB 849 VPPPTGDSGAPVPTGDSGAPVPTGDSGAPVPTGDSGAPVPTGDSGAPVPTGDSGAPVPT 907
 281 ---APPAPL--PGSSPPAP-----APPLPNSAPAPCPAPMGAPDPAPPLYS--- 327
 ||||| ||||| : ||||| ||||| : ||||| ||||| :
 DB 908 GDSGAPVPTGDSGAPVPTGDSGAPVPTGDSGAPVPTGDSGAPVPTGDSGAPVPTGDS 965
 QY 328 -SPAPAPCPVAPLAP 344
 :||| :||| :
 DB 966 GAPVPTGDSGAPVPTGDSGAPVPTGDSGAPVPTGDSGAPVPTGDSGAPVPTGDS 983

RESULT 10
 Q9LMQ1 PRELIMINARY; PRT; 1006 AA.
 AC Q9LMQ1;
 DT 01-OCN-2000 (TREMBlrel. 15, Created)
 DT 01-OCN-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-OCN-2000 (TREMBlrel. 15, Last annotation update)
 DE F7H2.17 PROTEIN.
 GN F7H2.17.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 CC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 CC Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RA Carrera A.J., Creasy T.H., Buell C.R., Town C.D., Nierman W.C.,

RA Liu S.X., Sakano H., Yu G., Etgu P., Lee J., Lenz C., Pham P.,
 RA Toriumi M., Chan A., Chung M., Goldsmith A., Liu A., Smith A.,
 RA Vaysberg M., Altafi H., Brooks S., Buehler E., Chao Q., Conn L.,
 RA Conway A.B., Hansen N.F., Johnson-Hopson C., Khan S., Kim C., Lam B.,
 RA Miranda M., Nguyen M., Palm C.J., Shih P., Southwick A., Davis R.W.,
 RA Ecker J.R., Federpsiel N.A., Theologis A.;
 RT "The sequence of BAC F7H2 from Arabidopsis thaliana chromosome 1";
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AC034256; AAF82153.1; -
 SO SEQUENCE 1006 AA; 103943 MW; C9FB49F9930C238D CRC64;

Query Match 27.3%; Score 605.5; DB 10; Length 1006;
 Best Local Similarity 33.3%; Pred. No. 2.9e-36;

Matches 176; Conservative 26; Mismatches 135; Indels 191; Gaps 29;

QY 4 PVPAPAP-----PSP--INPVPVPLPAAPRTLSPPVPAAPSPSTSLAAPLPDPMP 59
 ||||| ||||| : ||||| ||||| : ||||| ||||| :
 DB 99 PVPAP 152
 QY 60 PAIMSALEAPNPV-----PAPGNSAPAPMPPTPLPVPVPPGSGAPVPAV 111
 ||||| ||||| : ||||| ||||| : ||||| ||||| :
 DB 153 PP--SPL--VPSPPSP 208
 QY 112 PMPAPKRMPL-----PAPAPAPSPPTSLAVP--VP----- 143
 ||||| ||||| : ||||| ||||| : ||||| ||||| :
 DB 209 PVPPTGDETPVFSLPPLDDEFPPMPTLWLPDPVPAQTSAAEAFDQIPPLVITTEIE 268
 QY 144 ----- 146
 DB 269 NPNSHRRHDEKKGOLDRRNRVRSRRSRNGEAFSTRCDFEFCIFGTGGMNFPID 328
 QY 147 PVP-----PLVKKPSPVPVPPAPAPETP--NPAP-----AP-----LENSPP- 187
 ||||| ||||| : ||||| ||||| : ||||| ||||| :
 DB 329 PCQNPFLPPATLPPLPLPP--PPSLPVTCSPPPPPIYNGAPPPCVTCVOVSPPP 387
 QY 188 -----PVPVPPV--PVPPLTLNPVPPAPPAANTSNLRLP-----APPAPLK 231
 ||||| ||||| : ||||| ||||| : ||||| ||||| :
 DB 388 TPVPCSPPPPPPIPVPCPPSPSPPPPPQPCICVTAPAPPPQPCITVIAVASPPQ 447
 QY 232 P-----GPPAPMP-----PAPNS--PAAPSPSPSPVPVFPPTPPGP-----APPAPN 273
 ||||| ||||| : ||||| ||||| : ||||| ||||| :
 DB 448 PCITCVAAPEPPPPQPCITCIPAPASPPPPVPIPVFTPIFILPLPLPPLVLPSPSVT 507
 QY 274 SSPAP 314
 ||||| ||||| : ||||| ||||| : ||||| ||||| :
 DB 508 PSQLPLPPPSAPLPPPLSSLSPPPLPLVLSPPPLGGTVSQPPTMTPLPPLGGAP 567
 QY 315 GAPPPAPAPPLYS-----SPAP-----PACPVGAPLABLPISGRP 351
 ||||| ||||| : ||||| ||||| : ||||| ||||| :
 DB 568 GTTDSPPPLPILGSGAPGITSPPPLPILGSGAPGITSPPPLPILGSGAP 615

RESULT 11
 Q9ZQIO PRELIMINARY; PRT; 761 AA.
 AC Q9ZQIO;
 DT 01-MAY-1999 (TREMBlrel. 10, Created)
 DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
 DT 01-JUN-2000 (TREMBlrel. 14, Last annotation update)
 DE PUTATIVE PROLINE-RICH PROTEIN PRP2.
 GN F1K2.4.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 CC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 CC Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RA Lin X., Kaul S., Shea T.P., Fujii C.Y., Shen M., Vanaken S.E.,
 RA Barnstead M.E., Mason T.M., Bowman C.L., Ronning C.M., Benito M.,
 RA Carrera A.J., Creasy T.H., Buell C.R., Town C.D., Nierman W.C.,

DR PRINTS: PRO1217; PRICHEXTENS.
DR PRINTS: PRO1222; ATROPHIN.
SQ SEQUENCE 839 AA; 90243 MW; 0F0F0422D3D9789 CRC64;

Query Match 25.8%; Score 571.5; DB 10; Length 839;
Best Local Similarity 30.7%; Pred. No. 6.9e-34;
Matches 150; Conservative 36; Mismatches 127; Indels 175; Gaps 22;

QY 1 NSMPPVPAAPPPS-----PILNPPVPPVPLAAPTLSPPV 37
DB 370 NCLPAPRQORSGCGCAASLSLPVDCSGFCGRSTRPPVVPSPPTTSPGSG---SPPS 425
QY 38 PPAPPS-DISLAAPRLPDDPMPRAIMSALEARNPPVPPAPRGPSAPAPMPPTPLPP 96
DB 426 PISISPPITVSPPTTSPGSGP-----PSPSIVSP--PSTTPSGSPPTSPPTP 475
QY 97 VPFGSGARVPVAVP--PMPAPKMKRALPPAPAPSPPTSKLAVPPVPPVPPVPLPKM 155
DB 476 TPGGSPSSPTTPPGSGSPSSPTTPPGSGSPSSPTTPPGSGSPSSISPSPTIVPS 535
QY 156 PPSPPVPFPAPPEPTNP-----PAPAP-----PLENSPP-----PP--PVPPVP 196
DB 536 PPSPTTSGSPSSPSSTPSSPISPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 595
QY 197 VPPLTLNPPVPPAPPAAN-TSNSPLRPPAPAP----- 228
DB 596 SPPL--PPVPSPPVIGPPTSSP--PSTPTPGTLNPNHLLPQLSHLRHQLVHNHRL 650
QY 229 -----PL-----KRRPAP--- 237
DB 651 HILSRHRLRLRKHTIHRNHLHNPRMLQETALPLRLYLTCRHRMLASPPAPAYYY 710
QY 238 --PMPAPNSDAP-----SPSPVPVFPPTP-----GPPAPPEPSSPPA 278
DB 711 SSPQPPPPHSLPPTPTTYHTISPPPTTSHSPPOSHPPCLEYSPPPTVAYNPP 770
QY 279 PPAP--AAPLGP-----SPAPPA-----PPLNSPAPPGPAPGAPADPPAPPL 325
DB 771 PPSAHISPPSPVYYKNSPPPPAVHXSPPPPVHHSQPPPPRIEG-PLPPIPOIS 829
QY 326 YSSPPAP 333
DB 830 YASPPPP 837

RESULT 14
009083 PRELIMINARY; PRT; 580 AA.

DT 01-NOV-1996 (Tremblrel, 01, Created)
DT 01-NOV-1996 (Tremblrel, 01, Last sequence update)
DT 01-JUN-2000 (Tremblrel, 14, Last annotation update)
DE HYDROXYPROLINE-RICH GLYCOPROTEIN PRECUNSOR.
OS Phaseolus vulgaris (Kidney bean) (French bean).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
OC Fabales; Fabaceae; Papilionoideae; Phaseolus.
OX NCBI_TaxID=3885;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96061709; Pubmed=7480331;
RA Wycoff K.L., Powell P.A., Gonzales R.A., Corbin D.R., Lamb C.,
Dixon R.A.;
RT "Stress activation of a bean hydroxyproline-rich glycoprotein promoter
is superimposed on a pattern of tissue-specific developmental
expression.";
RL Plant Physiol. 109:41-52(1995).
RN [2]
RP SEQUENCE OF 5-229 FROM N.A.
RC TISSUE-LEAF AND STEM;
RX MEDLINE=86142825; Pubmed=3437892;
RA Corbin D.R., Sauer N., Lamb C.J.;

RT "Differential regulation of a hydroxyproline-rich glycoprotein gene
family in wounded and infected plants.";
RL Mol. Cell. Biol. 7:4337-4344(1987).

DR EMBL: U18791; AAA87902.1; -
DR EMBL: M18094; AAA3764.1; -
DR MENDEL: 16549; Phavu; 2930; 16549.

FT SIGNAL. 1 27 POTENTIAL.
FT CHAIN 28 580 HYDROXYPROLINE-RICH GLYCOPROTEIN.
SQ SEQUENCE 580 AA; 66316 MW; A97881835DABD4F CRC64;

Query Match 25.8%; Score 571; DB 10; Length 580;
Best Local Similarity 33.8%; Pred. No. 5.7e-34;
Matches 166; Conservative 29; Mismatches 152; Indels 144; Gaps 30;

QY 1 NSMPPV--PAP-----APSPINPPVPPV-----PP--LPAPRTLSPPVPP--APPS 43
DB 61 HSPPPPKHSPPPPPYYHNSPPPKHSPPPPPYYHNSPPPKHSPPPPPYYHNSPPPKHSP 120
QY 44 PISLAAPRLPDDPMPRAIMSALEAP--NPVPP-----APGPNAPAP-----MPPPPL 94
DB 121 PYYHNSPPPKHSPPPPPYYHNSPPPKHSPPPPPYYHNSPPPKHSPPPPPYYHNSPP 180
QY 95 PPVP--GSGARVPVAVP-----MPAPKRMBA-----LPAPAPSPPTSKLAVP 141
DB 181 SPPPPPYYHNSPPPKHSPPPPPYYHNSPPPKHSPPPPPYYHNSPPPKHSPPPPPYYHNS 240
QY 142 VPVPVPPVPLVKKMPSP--PVPPF-----PAEP-----ETPP--P 175
DB 241 PPKHSPPPPPYYHNSPPPKHSPPPPPYYHNSPPPKHSPPPPPYYHNSPPPKHSPPP 300
QY 176 APAPPLENSPPPPV-----PPV-----PVPPPLTLNPPVPA-----PPAAMTSN 217
DB 301 SPPPPPYYHNSPPPKHSPPPPPYYHNSPPPKHSPPPPPYYHNSPPPKHSPPPPPYY 360
QY 218 SPLRP--PAPAPPLK--PGP-----PAPPM-----PPAPNSAPAPSP--PSP 255
DB 361 SPPPPPYYHNSPPPKHSPPPPPYYHNSPPPKHSPPPPPYYHNSPPPKHSPPPPPYY 420
QY 256 PVVPF--PPPP-----GPPAPPEPNSPPAP--PAPAPPLGSPAP----- 295
DB 421 PPVYKYNSSPPPPYYKYSPPPPYYKYSPPPPYYKYSPPPPYYKYSPPPPYYKYS 480
QY 296 -----PAPPLNSPAPAPGPPAPGAPDPP-----APPLP-----YSSPPAPACPVGAP 341
DB 481 PPVYKYNSSPPPPYYKYSPPPPYYKYSPPPPYYKYSPPPPYYKYSPPPPYYKYS 540
QY 342 L-----APLP 347
DB 541 VYKYSPPPPV 551

RESULT 15
009082 PRELIMINARY; PRT; 388 AA.

DT 01-NOV-1996 (Tremblrel, 01, Created)
DT 01-NOV-1996 (Tremblrel, 01, Last sequence update)
DT 01-JUN-2000 (Tremblrel, 14, Last annotation update)
DE EXTENSIN CLASS I (CELL WALL HYDROXYPROLINE-RICH GLYCOPROTEIN) (HRCGP)
DE (TOM J-10).
GN TEGI.
OS Lycopersicon esculentum (Tomato).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I;
OC Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4081;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV, UC82B;
RX MEDLINE=92385769; Pubmed=1381233;
RA Zhou J., Rumeau D., Showalter A.M.;

RT *Isolation and characterization of two wound-regulated tomato extensin
RT genes.
RL Plant Mol. Biol. 20:5-17(1992).
CC -I- FUNCTION: HAS A SPECIALIZED STRUCTURAL FUNCTION, POSSIBLY IN THE
MECHANICAL PENETRATION OF THE CORTEX AND EPIDERMIS OF THE MAIN
ROOT. INVOLVED IN PLANT DEFENSE AGAINST INFECTION.
CC -I- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX.
CC -I- TISSUE SPECIFICITY: STEM.
CC -I- INDUCTION: BY WOUNDING AND PATHOGEN INFECTION.
CC -I- PTM: EXTENSINS CONTAIN A CHARACTERISTIC REPEAT OF THE PENTAPEPTIDE
SER-PRO(4). THE PROLINE RESIDUE IS HYDROXYLATED AND THEN
GLYCOSYLATED.
DR EMBL: M76670; AAA34163.1;
DR MENDEL: 16548; Lycos:2930;16548.
KM Repeat: Cell wall; Glycoprotein; Structural protein; Hydroxylation;
KM Multigene family.
FT DOMAIN 14 241 19 X 12 AA TANDEM REPEATS OF
S-P-P-P-P-S-P-K-Y-V-Y-K.
FT REPEAT 14 241 8 X 16 AA APPROXIMATE TANDEM REPEATS OF
S-P-P-P-P-Y-Y-Y-K-S-P-P-P-S-P.
FT REPEAT 14 25 1-1.
FT REPEAT 26 37 1-2.
FT REPEAT 38 49 1-3.
FT REPEAT 50 61 1-4.
FT REPEAT 62 73 1-5.
FT REPEAT 74 85 1-6.
FT REPEAT 86 97 1-7.
FT REPEAT 98 109 1-8.
FT REPEAT 110 121 1-9.
FT REPEAT 122 133 1-10.
FT REPEAT 134 145 1-11.
FT REPEAT 146 157 1-12.
FT REPEAT 158 169 1-13.
FT REPEAT 170 181 1-14.
FT REPEAT 182 193 1-15.
FT REPEAT 194 205 1-16.
FT REPEAT 206 217 1-17.
FT REPEAT 218 229 1-18.
FT REPEAT 230 241 1-19.
FT REPEAT 242 258 2-1.
FT REPEAT 259 274 2-2.
FT REPEAT 275 290 2-3.
FT REPEAT 291 306 2-4.
FT REPEAT 307 322 2-5.
FT REPEAT 323 338 2-6.
FT REPEAT 339 354 2-7.
FT REPEAT 355 370 2-8.
FT SEQUENCE 388 AA: 43227 MW: 9BA8B54AD8304647 CRC64;
Query Match 25.2%; Score 558; DB 10; Length 388;
Best Local Similarity 36.5%; Pred. No. 3.5e-33;
Matches 149; Conservative 27; Mismatches 140; Indels 92; Gaps 23;

OY 261 PTPGCP-PAPPEPNSPPAP-PAP-----AAPLPGSPAP-----PAPLPNSPAP 307
DB 285 PPPPSPPPPYYKSPPPSPSPPPYYKCPPPSPSPPPYYKSPPPYYKSPPPSPSP----- 341
OY 308 PGPPAMGAPDPPA-----PPLPYSSPPAPACPVGAPLAPLPIGSRP 351
DB 342 --PPYYHSPPPPVNSPPPPYYKSPPPYYKSPPP-----PYIYGSP 382
Search completed: January 5, 2001, 14:30:53
Job time: 345 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 5, 2001, 14:31:21 ; Search time 35.06 Seconds
(without alignments) 355.548 Million cell updates/sec

Title: US-09-461-774-10

Perfect score: 2216
Sequence: 1 NSMKVPVPAPPAPSPINPV.....NGAAAAAALAVAPAPAVKV 386

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 88757 seqs, 32294092 residues

Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	608	27.4	620	1	EXTN_TOBAC
2	511	23.1	817	1	VRP1_YEAST
3	505	22.8	5179	1	SP62_HUMAN
4	469.5	21.2	485	1	SP62_MOUSE
5	467	21.1	296	1	PRP3_MOUSE
6	450.5	20.3	1183	1	DRPL_RAT
7	445.5	20.1	346	1	PRF1_LYCES
8	439.5	19.8	426	1	EXLP_TOBAC
9	434.5	19.6	267	1	EXLP_MAIZE
10	432	19.5	1664	1	SLP1_CLOTM
11	421	19.0	464	1	SP62_HUMAN
12	419	18.9	1185	1	DRPL_HUMAN
13	415	18.7	434	1	TRX2_HUMAN
14	413	18.6	2715	1	TRX2_HUMAN
15	412	18.6	331	1	PRP1_HUMAN
16	409.5	18.5	1685	1	CAS4_HUMAN
17	408.5	18.4	1794	1	YAVI_SCHPO
18	408	18.4	283	1	EXTN_SORBI
19	408	18.4	487	1	EBN2_EBV
20	407.5	18.4	3164	1	TEGU_HSV11
21	402.5	18.2	1255	1	DIAL_MOUSE
22	398	18.0	2142	1	BAT2_HUMAN
23	391.5	17.7	1516	1	CAIH_HUMAN
24	388	17.5	865	1	CPN_DROME
25	388	17.5	1248	1	DIAL_MOUSE
26	387.5	17.5	1527	1	CAIH_MOUSE
27	386	17.4	261	1	PRP2_MOUSE
28	386	17.4	1255	1	MUC1_HUMAN
29	386	17.4	3421	1	TEGU_HSV1
30	384	17.3	633	1	LAI1_YEAST
31	382	17.2	1083	1	T2D3_HUMAN
32	378.5	17.1	1690	1	CA44_HUMAN
33	376	17.0	306	1	EXTN_DAVCA

34	374.5	16.9	1838	1	CA15_HUMAN
35	372.5	16.8	826	1	SSP2_PLAIO
36	367.5	16.6	680	1	CA1A_HUMAN
37	362	16.3	3149	1	TEGU_EBV
38	359	16.2	744	1	CA18_RABIT
39	357.5	16.1	1453	1	CA11_CHICK
40	356.5	16.1	296	1	CCOL_CARTEL
41	353	15.9	534	1	APG_ARATH
42	350.5	15.8	1638	1	BRM_DROME
43	348	15.7	371	1	PRP2_MEDTR
44	348	15.7	1157	1	SRA4_HUMAN
45	345.5	15.6	439	1	XP2_XENLA

ALIGNMENTS

```

RESULT 1
ID      EXTN_TOBAC      STANDARD:      PRT:      620 AA.
AC      P13983;
DT      01-JAN-1990 (Rel. 13, Created)
DT      01-JAN-1990 (Rel. 13, Last sequence update)
DT      01-MAR-1992 (Rel. 21, Last annotation update)
DE      EXTENSIN PRECURSOR (CELL WALL HYDROXYPROLINE-RICH GLYCOPROTEIN).
GN      HRCPRN3.
OS      Nicotiana tabacum (Common tobacco).
OC      Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC      Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I;
OC      Solanales; Solanaceae; Nicotiana.
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=CV, XANTHI; TISSUE=LEAF;
RX      MEDLINE=90128263; PubMed=2612909;
RA      Keller B., Lamb C.J.;
RT      "Specific expression of a novel cell wall hydroxyproline-rich
RL      glycoprotein gene in lateral root initiation.";
RL      Genes Dev. 3:1639-1646(1989).
CC      -1- FUNCTION: HAS A SPECIALIZED STRUCTURAL FUNCTION, POSSIBLY IN
CC      THE MECHANICAL PENETRATION OF THE CORTEX AND EPIDERMIS OF THE
CC      MAIN ROOT.
CC      -1- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX.
CC      -1- PTM: EXTENSINS CONTAIN A CHARACTERISTIC REPEAT OF THE PENTAPEPTIDE
CC      SER-PRO(4). THE PROLINE RESIDUE IS HYDROXYLATED AND THEN
CC      GLYCOSYLATED.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL: X13885; CAA32090.1; -.
DR      PIR: S06733; S06733.
KW      Repeat; Cell wall; Glycoprotein; Signal; Structural protein;
KW      Hydroxylation.
FT      SIGNAL          1          ?
FT      CHAIN           1          620
FT      REPEAT          70         73
FT      REPEAT          148        151
FT      REPEAT          229        242
FT      REPEAT          236        242
FT      REPEAT          205        620
FT      DOMAIN          499        600
FT      DOMAIN          620 AA; 65406 MM; 641DD227BAB28524 CRC64;
SQ      SEQUENCE

```

Query Match 27.4%; Score 608; DB 1; Length 620;
Best Local Similarity 37.3%; Pred. No. 8.2e-19;
Matches 166; Conservative 29; Mismatches 138; Indels 112; Gaps 29;


```

RESULT 7
PREF_LYCES STANDARD: PRT: 346 AA.
ID PREF_LYCES
AC 000451;
DT 01-OCR-1996 (Rel. 34, Created)
DT 01-OCR-1996 (Rel. 34, Last sequence update)
DT 01-OCR-1996 (Rel. 34, Last annotation update)
DE 36.4 KDA PROLINE-RICH PROTEIN.
CN TRPP-FL.
OS Lycopersicon esculentum (Tomato).
OS Euryalyta; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I;
CC Solanales; Solanaceae; Solanum.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. VENT CHERRY; TISSUE=FRUIT;
RX MEDLINE=92119262; PubMed=1731999;
SA Salts Y., Wachs R., Kenigsbuch D., Gruijssem W., Bary R.;
"DNA sequence of the tomato fruit expressed proline-rich protein gene
TRPP-FL reveals an intron within the 3 untranslated transcript.";
PL Plant Mol. Biol. 18:407-409(1992).
RN [2]
RP SEQUENCE OF 34-346 FROM N.A.
RC STRAIN-CV. ARAVA;
RX MEDLINE=91329722; PubMed=1868217;
SA Salts Y., Wachs R., Gruijssem W., Bary R.;
"Sequence coding for a novel proline-rich protein preferentially
RT expressed in young tomato fruit.";
PL Plant Mol. Biol. 17:149-150(1991).
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X61395; CAA43666.1; -
DR EMBL; X57076; CAA40361.1; -
DR HSSP; P24337; IHP.
DR INTERPRO; IPR000528; -
DR PFAM; PF00279; LTP; 1.
DR SEQUENCE 346 AA; 36375 MM; 604E58452FEF16D CRC64;

Query Match 20.1%; Score 445.5; DB 1; Length 346;
Best Local Similarity 34.6%; Pred. No. 1.8e-12;
Matches 119; Conservative 23; Mismatches 111; Indels 91; Gaps 19;

OY 7 PAPRPPSPINPVPPVPLPAPRPLTSLPPVPPAPSPSLAAPLPDPPPAPPLMSL 66
DB 8 PUCVPSPS--TPKHPLPCKVKPSTOPPHVKP-----PSTPKHKKDPPHV---- 51
OY 67 EAPNPVPVPPAPGNSAPAPRPPPLPP--VPSCGAPRVAVPMPAPKRRAPALP 125
DB 52 ---KPPSPKQP-----FYVKPPTTPKHPHVKP-----PSTPKHKKDPPKCPP 95
OY 126 APPAPSPPTSWLAVPVPPVPPVPLPVKMPSPVPPEPPAPPEPTPNPAPAPLENS 185
DB 96 SHHGKRPVKKRPHVBRPI--VHPRPVSPSTPKRPTK--PFTPKRPS-PDPPIVS- 149
OY 186 PPPPVPPVPPVPLTLNPVPPAPPAANTSPLRPAAPRPLKPGRRAPMPAPNS 245
DB 150 -PPIVYPITTPPI-VHPVTPKPPS-----PPIPI-----VSPPI----- 184
OY 246 PAAPSPSPSPV---PVPPTPP--GPAPAPPN--SSPPAPAPAPALPGSPAPAPAP 298
DB 185 -VYPITTPPVSPVSPITTPPIVSPFPVNPVVIIPPVYVSPVY-----TPPIVTPP 238
OY 299 PLPNSPAAPGPPAMPGADPPAPPLPYSSPPAPAPACVPAGAPL 342

```

```

Db 239 PTP-----CPPPPPPAIPSPAPQPTCPIDAKL 268

RESULT 8
EXLP_TOBAC STANDARD: PRT: 426 AA.
ID EXLP_TOBAC
AC 003211;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE 15-JUL-1999 (Rel. 38, Last annotation update)
DE PISTIL-SPECIFIC EXTENSIN-LIKE PROTEIN PRECURSOR (PELP).
OS Nicotiana tabacum (Common tobacco).
OS Euryalyta; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I;
CC Solanales; Solanaceae; Nicotiana.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. PETITE HAVANA; TISSUE=PISTIL;
RX MEDLINE=93005740; PubMed=1392607;
SA Goldman M.H., Pezzotti M., Seurinck J., Mariani C.;
"Developmental expression of tobacco pistil-specific genes encoding
RT novel extensin-like proteins.";
PL Plant Cell 4:1041-1051(1992).
RL -1- TISSUE SPECIFICITY: PISTIL (STIGMA AND STYLE TISSUE).
CC -1- DEVELOPMENTAL STAGE: EXPRESSION BEGINS IN FLORAL BUDS AFTER
CC PISTIL DIFFERENTIATION AND LEVELS GRADUALLY INCREASE DURING
CC FLOWER DEVELOPMENT TOWARD ANTHESIS. LEVELS GRADUALLY DECREASE
CC AFTER POLLINATION AND ARE ABSENT BY THE SIXTH DAY AFTER
CC POLLINATION.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; Z14019; CAA78397.1; -
DR PIR; J01696; J01696.
DR INTERPRO; IPR000419; -
DR PFAM; PF01190; Pollen_Ole_e_I; 1.
KW Structural protein; Signal; Repeat; Glycoprotein.
FT SIGNAL 1 23
FT CHAIN 24 426 PISTIL-SPECIFIC EXTENSIN-LIKE PROTEIN.
FT DOMAIN 69 182 4 x 5 AA REPEATS OF S-P(4).
FT REPEAT 69 73 1.
FT REPEAT 76 80 2.
FT REPEAT 83 87 3.
FT REPEAT 178 182 4.
FT CARBOHYD 310 310 N-LINKED (GLCNAC...) (POTENTIAL).
SO SEQUENCE 426 AA; 44278 MM; 51A495C94017812 CRC64;

Query Match 19.8%; Score 439.5; DB 1; Length 426;
Best Local Similarity 36.4%; Pred. No. 3.5e-12;
Matches 128; Conservative 24; Mismatches 85; Indels 115; Gaps 25;

OY 51 PLPDPMPAPLWALPAPNPVPVPPAPGPNAPAPMPPTPLPVPVPGGAPRPVAV 110
DB 35 PLPFD-----WPAELPLPDI-----SSPFGPTFVLPP-----PSPLSP 70
OY 111 PMPAPKRMALPAPAPSPPTSWLAVPVPP-----VPLPVKMPSPVPVPPF 165
DB 71 PPPSPSP-----PPSPSPSPST-----IPLPPTGTGLPL-----PGSKLPDFA 113
OY 166 PAEPETPNPAPAPALNSPPPPVPPVPPVPLTLNPVPPAPPAANTSPL-REPA 224
DB 114 GLPLPLRN--LPDVPTTGGPP-----VQPKRPS-----SPLVKRPP 150
OY 225 PPAPLKGPP-----ADMPAPNSPAAPSPSPVPVPPPTPPGPPAP--PEPNSPPA 278

```

DB 151 PPPSPCKSPSPDOSAKOPQPP-----PAKOPSPPPPPPPV-----KAPSPSPAKQPP 199
 QY 279 PPAPPAAPLPGPS--PPA---PPAPPLPNSPAPPGPP--AMP-----GAPPPA--PPL--P 325
 DB 200 PPPPVKAPSPSPATQPTKQPPPPPRAKKSPLLPPPPVAVPPVMPSPSPAEPPIAP 259
 QY 326 YSSPPA-PPACPVGAP--LAPPLISGRP-----SNSWGVFTMLS 363
 DB 260 FPPPPANPLIPRRRPPAPVVKPLPLPLGKPPPIVSGIYCKSCSNY--GVPTLLN 310

RESULT 9
 EXTN_MAIZE
 ID EXTN_MAIZE STANDARD: PRT: 267 AA.
 AC P14918:
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE EXTENSIN PRECURSOR (PROLINE-RICH GLYCOPROTEIN).
 HGRP.
 Zea mays (Maize).
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 Magnoliophyta; Liliopsida; Poales; Poaceae; Zea.
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. W64A, AND CV. E41;
 RA Stiefel V., Perez-Grau L., Albericio F., Giral E., Ruiz-Avila L.,
 RA Ludevid M.D., Puigdomenech P.;
 RT "Molecular cloning of cDNAs encoding a putative cell wall protein from
 RT Zea mays and immunological identification of related polypeptides";
 RL Plant Mol. Biol. 11:483-493(1988).
 CC -1- FUNCTION: STRUCTURAL COMPONENT IN PRIMARY CELL WALL.
 CC -1- TISSUE SPECIFICITY: MAINLY IN THE COLEOPTILE NODE AND ROOT TIP.
 CC -1- PTM: EXTENSIN CONTAIN A CHARACTERISTIC REPEAT OF THE PENTAPEPTIDE
 CC SER-PRO(4). THE PROLINE RESIDUE IS HYDROXYLATED AND THEN
 CC GLYCOSYLATED.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X13499; CAA31854.1; -
 DR EMBL: X13506; CAA31860.1; -
 DR EMBL: M36912; AAA33455.1; -
 DR EMBL: M36913; AAA33456.1; -
 DR EMBL: M36914; AAA33457.1; -
 DR PIR: S08314; S08314.
 DR MAI2EDB: 17152; -
 DR INTERPRO: IPR002965; -
 DR PRINTS: PRO1217; PRICHTENS.
 KW Repeat: Cell wall; Glycoprotein; Signal; Structural protein;
 KM Hydroxylation.
 FT SIGNAL 1 ?
 FT CHAIN 267
 FT DOMAIN 18 253
 FT REPEAT 18 33
 FT REPEAT 34 54
 FT REPEAT 55 70
 FT REPEAT 71 91
 FT REPEAT 92 107
 FT REPEAT 108 128
 FT REPEAT 129 144
 FT REPEAT 145 160
 FT REPEAT 161 179
 FT REPEAT 180 195
 FT REPEAT 196 211
 FT REPEAT 212 232
 FT REPEAT 233 253
 FT DOMAIN 261 265
 EXTENSIN REPETITIVE ELEMENT.

FT VARIANT 245 245 MISSING (IN STRAIN E41).
 FT VARIANT 261 261 S -> C (IN STRAIN E41).
 SQ SEQUENCE 267 AA; 28349 MM; A6F406445FEECB CRC64;

Query Match 19.6%; Score 434.5; DB 1; Length 267;
 Best local Similarity 36.8%; Pred. No. 4.1e-12;
 Matches 107; Conservative 15; Mismatches 114; Indels 55; Gaps 17;

QY 32 TISPPV-PPAPSPDISLADPLDPDPMPIAISALEAPNPVP-PAPGNSAPAPWP 89
 DB 15 SLTPPTPTSPKPPPTPKPPPTPTSPKPP-----ASKPPPKPPPTPTSPKPPPP 67
 QY 90 -PTPPL---PPVPGSAGAPVAVPMPAPKRMALPAPAPSPSPSWLAVVPPV 145
 DB 68 KPTPTPTTSPKPPATKPPPTPTPTTSPK-----PTPT--KPTPT-----Y 111
 QY 146 PPVPLPVKMPSPPPVPPPPAPETPNPAP-PAPPLSNSPPPPVPPVPPVPLTLP 204
 DB 112 TFSKPPATKPPPT--KPTPTTSPKPPPTPKPTPTPTTSPKPPPT--KPTPTTSP 167
 QY 205 PVPAPPAANTSNPLRPPAPAPPLKGGPADMPAPNSAPAPSPSPVPPPTPP 264
 DB 168 PKPPTHT-----PKPTPTTSPKPP--TPKPTPTTSPS--PKPTPK--PTPP 213
 QY 265 G-PPAPPEPNSPPAPPAAPLPGSPPA-----PPAPPLPNSPAPAP 308
 DB 214 TTPSPKPPATKPPPTKPPPTTPTTPKPPATKPPPTTTPPVSHTPSPPP 264

RESULT 10
 ID SLPL_CLOTM STANDARD: PRT: 1664 AA.
 AC 006852;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 01-FEB-1996 (Rel. 33, Last annotation update)
 DE CELL SURFACE GLYCOPROTEIN 1 PRECURSOR (OUTER LAYER PROTEIN B) (S-LAYER
 DE PROTEIN 1).
 GN OLPB.
 OS Clostridium thermocellum.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
 OC Clostridium.
 CC [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NCIB 10682;
 RX MEDLINE=93209931; PubMed=8458832;
 RA Fujino T., Beguin P., Aubert J.-P.;
 RT "Organization of a Clostridium thermocellum gene cluster encoding the
 RT cellulosomal scaffolding protein ClpA and a protein possibly involved
 RT in attachment of the cellulosome to the cell surface.";
 RL J. Bacteriol. 175:1891-1899(1993).
 CC -1- SUBUNIT: ASSEMBLED INTO MONO-LAYERED CRYSTALLINE ARRAYS.
 CC -1- SUBCELLULAR LOCATION: CELL WALL.
 CC -1- SIMILARITY: CONTAINS 4 S-LAYER HOMOLOGY (SLH) DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X67506; CAA47841.1; -
 DR INTERPRO: IPR001119; -
 DR PRAM: PFO0395; SLH: 3.
 DR PROSITE: PS01072; SLH_DOMAIN: 2.
 KW Cell wall; S-layer; Signal; Repeat.
 FT SIGNAL 1 28
 FT CHAIN 29 1664
 FT DOMAIN 36 763
 FT REPEAT 36 191
 POTENTIAL.
 4 X 156 AA APPROXIMATE REPEATS.
 1.

FT REPEAT 207 363 2.
 FT REPEAT 409 565 3.
 FT REPEAT 607 763 4.
 FT DOMAIN 771 1377 APPROXIMATE TANDEM REPEATS OF
 T-P-S-D-E-P.
 FT DOMAIN 1378 1449 GLY/PRO/SER/THR-RICH.
 FT DOMAIN 1453 1494 SLH 1 (INCOMPLETE).
 FT DOMAIN 1495 1565 SLH 2.
 FT DOMAIN 1566 1625 SLH 3.
 FT DOMAIN 1626 1646 SLH 4 (INCOMPLETE).
 SQ SEQUENCE 1664 AA; 178194 MM; 5F396695BA9FE74B CRC64;

Query Match 19.5%; Score 432; DB 1; Length 1664;
 Best Local Similarity 28.0%; Pred. No. 1.6e-11;
 Matches 130; Conservative 28; Mismatches 205; Indels 102; Gaps 21;

QY 4 PVP---PAPAPSPINPVV---PVPLPA-----APRTLSPVPAPAPSPISLAAP-- 50
 972 PTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSD 1031
 51 -PLPPDPMPAPMSALEAPMPVPPAPAPG---PNSAPAPMPPTPLPVPVPGSGAPR 105
 1032 EPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPT 1087
 QY 106 PVAPVPPMPAPKIMPALPAP-----PAPSPPTSLAVPVPPVPPVPLPYKMP 156
 1088 PTPSETPEEPIPTDTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPT 1146
 QY 157 PSPVP-----PFPAPAP---ETPNPAP-----PAPPLENS 185
 1147 SDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPT 1206
 QY 186 PPPPVPPVPPVPLTLNPPV-----PAPPAANTNSPLRPPAPAPPLKGGPPAPR 238
 1207 PPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPT 1266
 QY 239 MPAP-NSPAPSPSPVPV--PPTPPGPPAP---DEPNSSP-----PAPPAAPALP 288
 1267 EEPITPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPT 1366
 QY 289 GPSP---PAPAPPLPN---SPAAPGPPAPGAPPPAP---PLPYSSP--PAPPACV 337
 1327 TDPSPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPS 1386
 QY 338 PGAPLAPLPISGRPSNMGVFTMLSRPSNGAANAASALAVAPAR 382
 1387 PTSTPTSGSGSGSGS-----GGGGGGGGGTPTSPTR 1420

RESULT 11
 SP62_HUMAN STANDARD; PRT; 464 AA.
 AC 015428;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DE 01-OCT-2000 (Rel. 40, Last annotation update)
 DE SPLICEOSOME ASSOCIATED PROTEIN 62 (SAP 62) (SPLICING FACTOR 3A SUBUNIT 2) (SF3A62).
 GN SF3A2 OR SAP62.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-94023929; PubMed-8211113;
 RA Bennett M., Reed R.;
 RT "Correspondence between a mammalian spliceosome component and an essential yeast splicing factor.";
 RL Science 262:105-108(1993).
 CC -1- FUNCTION: SUBUNIT OF THE SF3A COMPLEX REQUIRED FOR 'A' COMPLEX ASSEMBLY FORMED BY THE STABLE BINDING OF U2 SNRNP TO THE BRANCHPOINT SEQUENCE (BPS) IN PRE-mRNA. SEQUENCE INDEPENDENT

CC BINDING OF SF3A/SF3B COMPLEX UPSTREAM OF THE BRANCH SITE IS
 CC ESSENTIAL. IT MAY ANCHOR U2 SNRNP TO THE PRE-mRNA.
 CC -1- SUBUNIT: SF3A IS COMPOSED OF SAPS 61, 62 AND 114 (SF3A60, SF3A60
 CC AND SF3A120 RESPECTIVELY).
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
 CC -1- SIMILARITY: TO YEAST PRP1.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: L21990; AAA60301.1; -.
 DR MIM: 600796; -.
 KW mRNA processing; mRNA splicing; Nuclear protein; Repeat.
 FT DOMAIN 233 236 POLY-PRO.
 FT DOMAIN 252 256 POLY-PRO.
 FT DOMAIN 458 462 POLY-PRO.
 SQ SEQUENCE 464 AA; 49196 MM; 5D4C7B044AE81C71 CRC64;

Query Match 19.0%; Score 421; DB 1; Length 464;
 Best Local Similarity 39.4%; Pred. No. 2e-11;
 Matches 119; Conservative 12; Mismatches 101; Indels 70; Gaps 19;

QY 9 PPAPPS-PINPVPPVPPVPLAARTLSPPVPPAPSPISLAAPPLPD---PPMPAIW 63
 217 PPAPPSLPAGPPGVKRRPP---LMNGCLPRRPLRESL--PPPPGGLPLRPPM--- 265
 QY 64 SALEAPNPPVAPRGNSAPAPRMPRTPLPV--PRGSGAPRVVAPVPPAPKMPRA 122
 266 -----PTGRAPSGPPGRRQLRRAPGVNPPAVVNPAPSGVNPAPGV--NPAP---GV 315
 QY 123 LPPAPAPSPPTSLAVPVPPV--PPLVVKMPSPVPVPPPPAPRABETRNPPAPR 181
 316 HPAPAGV--HPRTSGVH--PRAPGVNPPAPGVNPPAPGV-----HPRPPGV-- 357
 QY 182 LENSPPPPVPPVPPVPLTLNPPV---PAPPAANTNSPLRPPAPRPLKGGPPAPR 238
 358 -----HPRAPGVNPPSAGVNHQAPGVNPPAPAVNPPQAPGVNPPAPGAMQAPG--VNP 409
 QY 239 MPAP-NSPAPSPSPVPVPPVPPPPG--PAPRPPNSPPAPRPPAPRPLPGSPRA 294
 410 QPQVNPASGVNPPQPG---VNPSPGVNPPPTMPPLRPPV-----SEGGMTPR 460
 QY 295 PP 296
 Db 461 PP 462

RESULT 12
 DRPL_HUMAN STANDARD; PRT; 1185 AA.
 ID DRPL_HUMAN
 AC P54259;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DE 15-JUL-1999 (Rel. 38, Last annotation update)
 DE ATROPHIN-1 (DENTATORUBRAL-PALLIDOLYSIAN ATROPHY PROTEIN).
 GN DRPLA.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-CEREBELLUM, AND BRAIN;
 RX MEDLINE-95144175; PubMed-7842016;
 RA Nagafuchi S., Yamagisawa H., Ohsaki E., Shitayama T., Tadokoro K., Inoue T., Yamada M.;
 RT "Structure and expression of the gene responsible for the triplet repeat disorder, dentatorubral and pallidolysian atrophy (DRPLA).";

RL Nat. Genet. 8:177-182(1994).
 RN [2].
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96262314; PubMed=8965642;
 RA Margolis R.L., Li S.-H., Young W.S., Wagster M.V., Stine O.C.,
 RA Kidwai A.S., Ashworth R.G., Ross C.A.;
 RT "DRPLA gene (atrophin-1) sequence and mRNA expression in human
 brain";
 RN Brain Res. Mol. Brain Res. 36:219-226(1996).
 RN [3].
 RP SEQUENCE OF 470-725 FROM N.A.
 RC TISSUE=BRAIN CORTEX;
 RX MEDLINE=93315145; PubMed=8325628;
 RA Li S.-H., McInnis M.G., Margolis R.L., Antonarakis S.E., Ross C.A.;
 RT "Novel triplet repeat containing genes in human brain: cloning,
 expression, and length polymorphisms";
 RL Genomics 16:572-579(1993).
 CC -1- TISSUE SPECIFICITY: THE LEVELS ARE RELATIVELY HIGH IN THE BRAIN,
 CC OVARY, TESTIS AND PROSTATE. LOWER LEVELS ARE DETECTED IN THE
 CC LIVER, THYMUS AND LEUKOCYTES.
 CC -1- POLYMORPHISM: THE POLY-GLN REGION OF DRPLA IS HIGHLY POLYMORPHIC
 CC (7 TO 23 REPEATS) IN THE NORMAL POPULATION AND IS EXPANDED TO
 CC ABOUT 49-75 REPEATS IN DRPLA PATIENTS. LONGER EXPANSIONS RESULT IN
 CC EARLIER ONSET AND MORE SEVERE CLINICAL MANIFESTATIONS OF THE
 CC DISEASE.
 CC -1- DISEASE: DEFECTS IN DRPLA ARE THE CAUSE OF DENTATORUBRAL-
 CC PALCIDOLYSIAN ATROPHY, AN AUTOSOMAL DOMINANT NEURODEGENERATIVE
 CC DISORDER CHARACTERIZED BY A LOSS OF NEURONS IN THE DENTATE
 CC NUCLEUS, RUBRUM, GLOBUS PALLIDUS AND LUTS/BODY. CLINICAL FEATURES
 CC ARE MYOCLONUS EPILEPSY, DEMENTIA, AND CEREBELLAR ATAXIA. ONSET OF
 CC THE DISEASE OCCURS USUALLY IN THE SECOND DECADE OF LIFE AND DEATH
 CC IN THE FOURTH.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation --
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; D31840; BAA06626.1; -;
 DR EMBL; U23851; AAB50276.1; -;
 DR EMBL; L10377; -; NOT_ANNOTATED_CDS.
 DR HSSP; P00651; ILFA.
 DR MIM; 125370; -;
 DR INTERPRO; IPR002951; -;
 DR PRINTS; PR01222; ATROPHIN.
 RN Triplet repeat expansion; Polymorphism.
 FT DOMAIN 73 82 SER/GLU-RICH (MIXED CHANGE).
 FT DOMAIN 302 305 POLY-PRO.
 FT DOMAIN 376 382 POLY-SER.
 FT DOMAIN 386 397 POLY-SER.
 FT DOMAIN 442 447 POLY-PRO.
 FT DOMAIN 479 483 POLY-HIS.
 FT DOMAIN 484 497 POLY-GLN.
 FT DOMAIN 504 507 POLY-PRO.
 FT DOMAIN 564 574 POLY-SER.
 FT DOMAIN 704 707 POLY-PRO.
 FT DOMAIN 802 815 ARG/GLU-RICH (MIXED CHANGE).
 FT DOMAIN 816 827 ARG/GLU-RICH (MIXED CHANGE).
 FT DOMAIN 925 934 ARG/GLU-RICH (MIXED CHANGE).
 FT CONFLICT 94 94 MISSING (IN REF. 2).
 FT CONFLICT 333 333 Y -> H (IN REF. 2).
 FT CONFLICT 339 339 M -> I (IN REF. 2).
 FT CONFLICT 541 541 P -> T (IN REF. 3).
 FT CONFLICT 1028 1028 G -> A (IN REF. 2).
 RN SEQUENCE 1185 AA; 124785 MW; 56C306267331C005 CRC64;

Query Match	18.9%;	Score 419;	DB 1;	Length 1185;
Best Local Similarity	26.3%;	Pred. No. 4.4e-11;		
Matches 161;	Conservative 37;	Mismatches 164;	Indels 250;	Gaps 30

OY	7	PAPRPPSPINPVRV-----PLRAPRT-----LSPP-----VPAPPS	43
Dd	155	PAPRPRLPRLPFRSQPRDSTPROREASEREPVSVTPTGYAAMERPYSRMFOABRCAR	214
OY	44	PIS-----LAAPPL-----PPDPMPRAIWSALEAVNPP	72
Dd	215	RHQLYRGGTGVLGRPMRGCGGAASSVGKNGCKOHRPTTPTISVSSGASGAR-PT	273
OY	73	VPRAPP--GNPAARPP-----MRPRLPV-----PPGSAPRRVVNVP--	112
Dd	274	KRPTRVGGGNLPSRAPRANFHTVTPNLRRPALRPLNNASAPSGLGA-OPLRGLHLPSP	332
OY	113	-----MPRARMPALPRAP-----PAP-----SPTT ₁₃₅	
Dd	333	YAMGGMGGLGPRPKGRTLAPSHSLRPASSAPARPRMRRPYSSSSSSAAASSSSSS	392
OY	136	SMLAVPVPRVPRVRLPYKMCRPVRPPRRPARERTMPRPARADLENSRPPRV--	192
Dd	393	SSASAPRFASQALPGYPSFR--PRTSLSYSNQPKYQPSLRSGAWMSQGRRPPRYGR	450
OY	193	-----EVPVY-----PRLTN-----PPVPPA-209	
Dd	451	LLANSMNPDRPFRPSTGAOSTAHNVSTNHNNHQOOOQQOOQHNGNSGRRPCAF	510
OY	210	-----PRANTSNPLRPPRARPRYLK-----GPRAPMRPAANS245	
Dd	511	PHLEGSSHHNHPAMSPLGSLR-PYRCPGANLRRHSQVYSQAQDRNGPVPSSSNS	569
OY	246	-----PAAPRSP-----PSPVPVVFPT-----PPG-----PP267	
Dd	570	SSSTQSIGYRCHNRSPRSQGRAGYRFPFRPVTVTTSSATLSIVITVAASSPAGYKTASP	629
OY	268	APERPNSSPRADARPAPRLGPPSPAPRA-----PP--LPNSPAPRPG-----PPAMP314	
Dd	630	GPPRYKRAPSPGAKTKTTPRCYKGPSPSPFRGTGNPRGNSPRACGTFFKGSPRYGR	689
OY	315	CARDPRAPRLPSSPPRAPACVPCAPRALPRLISRPSNSWGVFTMLSRPNGAAAAAS	374
Dd	690	G-LPLPPAGSGLLSLRPPRAPASGPPLSAQIKQEPAEY-----ETPESPVPAPRS741	
OY	375	ALAYAPAPAVKY386	
Dd	742	-----PSPRPKV748	

RESULT 13

ID	NO75.LUPLU	STANDARD:	PRT:	434 AA.
AC	006841:			
DT	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	01-NOV-1997 (Rel. 35, Last annotation update)			
DE	EARLY NODULIN 75 PROTEIN (N-75) (NGM-75) (FRAGMENT). ENOD2.			
OS	Lupinus luteus (Yellow lupine).			
OC	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;			
OC	Magnoliopsida; eudicotyledons; core eudicots; Rosidae; eurosids I;			
OC	Fabales; Fabaceae; Papilionoideae; Lupinus.			
RA	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=ROOT NODES;			
RX	MEDLINE=91355883; PubMed=2103455;			
RT	Szczeglowski K., Legocki A.B.,			
RT	"Isolation and nucleotide sequence of cDNA clone encoding nodule-			
RT	specific (hydroxy)proline-rich protein LENOD2 from yellow lupin.";			
RL	Plant Mol. Biol. 15:361-363(1990).			
CC	-1 FUNCTION: INVOLVED IN EARLY STAGES OF ROOT NODULE DEVELOPMENT.			
CC	-1 SIMILARITY: TO OTHER PLANTS N-75.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			

FT	DOMAIN	2586	2715	SET DOMAIN.
FT	DOMAIN	26	37	POLY-GLY.
FT	DOMAIN	248	255	POLY-PRO.
FT	DOMAIN	362	398	ASP/GLU- RICH (ACIDIC).
FT	DOMAIN	402	771	PRO-RICH.
FT	DOMAIN	808	812	POLY-GLN.
FT	DOMAIN	1963	1970	POLY-PRO.
FT	DOMAIN	2251	2259	POLY-PRO.
FT	VARSPLIC	532	582	VARSSRSVRIKTPRRMEDPPKPKVEVSPYLRRPITTSPP VOEAPAPVS -> PLISGLLPMTLQISLIGMAPTTS ACIDSPMSPLLRCPPLTGLQ (IN ISIFORM TRUNCATED).
FT				
FT				
FT				
FT	VARSPLIC	583	2715	MISSING (IN ISIFORM TRUNCATED).
FT	CONFLICT	834	834	K -> E (IN REF. 5).
FT	CONFLICT	941	941	S -> Y (IN REF. 5).
FT	CONFLICT	1317	1317	E -> Q (IN REF. 5).
FT	CONFLICT	1362	1362	H -> Y (IN REF. 5).
FT	CONFLICT	1438	1438	D -> N (IN REF. 5).
FT	CONFLICT	2622	2622	D -> H (IN REF. 5).
FT	SEQUENCE	2715	AA: 293511	MM: C0615B941BBEB7BF CRC64:

[illegible]

RP	SEQUENCE FROM N.A.
RX	MEDLINE=65289325; PubMed=2993301;
RT	Maeda N., Kim H.-S., Azen E.A., Smithies O.;
RT	"differential RNA splicing and post-translational cleavages in the
RT	human salivary proline-rich protein gene system.";
RL	J. Biol. Chem. 260:11123-11130(1985).
RN	[2]
RP	SEQUENCE OF 214-331.
RX	MEDLINE=86243355; PubMed=3521730;
RA	Kaufman D., Hofmann T., Bennick A., Keller P.;
RT	"Basic proline-rich proteins from human parotid saliva: complete
RT	covalent structures of proteins IB-1 and IB-6.";
RL	Biochemistry 25:2387-2392(1986).
RN	[3]
RP	SEQUENCE OF 276-331.
RX	MEDLINE=84161824; PubMed=6671974;
RA	Saitoh E., Isemura S., Sanada K.;
RT	"Further fractionation of basic proline-rich peptides from human
RT	parotid saliva and complete amino acid sequence of basic proline-rich
RT	peptide P-H.";
RL	J. Biochem. 94:1991-1997(1983).
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC	or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL, K03204; AAA60185.1; -
DR	EMBL, K03205; AAA60186.1; -
DR	EMBL, K03206; AAA60187.1; -
DR	PIR, A03291; PIHUB6.
DR	PIR, C25372; C25372.
DM	MIM; 168730; -
KW	Repeat; Parotid gland; Multigene family; Saliva; Signal.
FT	SIGNAL
FT	CHAIN
FT	CHAIN
FT	VARIANT
FT	VARIANT
FT	CONFLICT
FT	SEQUENCE
SO	
Query Match	18.6%; Score 412; DB 1; Length 331;
Best Local Similarity	34.2%; Pred. No. 3,8e-11;
Matches 111; Conservative 10; Mismatches 140; Indels 64; Gaps 15;	
OY	65 ALAENPPVPAAPCGPNSADAPMPTPLLPVPVPGSGAPR-PVPAVPMMPAPKMPA 122
D6	30 SLIAGNPQ-GSPDGGCKPQGPPPPGPKPQGPPGCKPQGPPPPGPKPQGPPGDKSR 88
OY	123 LPPAPAPPSPPTSLAVLVPPVPPVLVVKMPSVPVPPPEAEETPNPAPAPPL 182
D6	89 SPRSPGPKRGD-----PPOGGNOGPpppppgkpkqgpppoggnkkgpgr- 133
OY	183 ENSPPPVPPVPPVPLTLNPPVPAPPAANTSNPLRPPAPAPPLPKR-GPP-----A 236
D6	134 ---PPGKPGGR-PPOGGDKSQSPRRPGKRQGGPPGCGNOGGRPPPPGKPKQGGPPQGGNR 189
OY	237 PPMPPAPNSPAAP-----SPSPPYVVFETPRGPAPPEPNSSPPAPAPAPLPG 289
D6	190 PQGPPPPGKPKQGGPPQGGDKSKSPQSPP---GKPQGR--PPOGGNOGPPPPPPGKPK-QG 242
OY	290 PSPPA---PPAPPLLNSPAAP-----GPPAMPGAR-----DPAPPLPYSSP 329
D6	243 PPPGCGNKPKQGGPPPPGKPKQGGAPGGSKSQSAKARPPGKPKQGGPQEGNNPPPPACGN 302
OY	330 PAPPACPVGAPLAP-LPISGRPS 352

Db 303 PQQPAPPAQQPQGP RRRPPQGGRRPS 327

Search completed: January 5, 2001, 14:31:28
Job time: 359 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 5, 2001, 14:28:42 ; Search time 67.42 Seconds
(without alignments)
388.752 Million cell updates/sec

Title: US-09-461-774-10
Perfect score: 2216
Sequence: 1 NSMPVPAPAPAPPPINPV.....NGAAAASALAVAPAPAVKV 386

Scoring table:
BLOSUM62
Gapop 10.0 , Gapect 0.5

Searched: 195891 seqs, 67900655 residues
Total number of hits satisfying chosen parameters: 195891

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	687	31.0	760	2	T06291	extensin homolog T
2	670.5	30.3	1188	2	S49915	extensin-like prot
3	659	29.7	489	2	T11622	extensin class 1 p
4	608	27.4	620	2	S06733	hydroxyproline-ric
5	608	27.4	1585	2	T31611	hypothetical prote
6	571.5	25.8	839	2	T04859	extensin homolog F
7	571	25.8	580	2	T10863	extensin precursor
8	558	25.2	388	2	S25298	extensin (clone T0
9	555.5	25.1	432	2	T016782	extensin - soybean
10	542	24.5	786	2	T01456	extensin homolog F
11	533	24.1	416	2	TJ0465	extensin precursor
12	520	23.5	280	2	T11671	extensin-like prot
13	516.5	23.3	376	2	S71558	probable cell wall
14	511	23.1	817	2	S51342	verprolin - yeast
15	505	22.8	1611	2	T38236	hypothetical prote
16	505	22.8	3020	2	A43932	mucin 2 precursor
17	504.5	22.8	491	2	T07598	proline-rich prote
18	502	22.7	330	2	T05717	probable extensin
19	502	22.7	369	2	S20500	hydroxyproline-ric
20	502	22.7	951	2	T47617	extensin-like prot
21	499.5	22.5	464	2	S22697	extensin - Volvox
22	497.5	22.5	733	2	T47618	extensin-like prot
23	493.5	22.3	350	2	S22456	hydroxyproline-ric
24	493	22.2	589	2	T29299	hypothetical prote
25	487.5	22.0	707	2	T14195	extensin homolog T
26	484.5	21.9	322	2	S25299	extensin precursor
27	471	21.3	392	1	PIH086	salivary proline-r
28	470.5	21.2	839	2	F75518	hypothetical prote
29	467.5	21.1	328	2	J00985	hydroxyproline-ric

30	464.5	21.0	513	2	T14194	extensin homolog T
31	463	20.9	317	2	A28996	proline-rich prote
32	462.5	20.9	368	2	C29356	hydroxyproline-ric
33	456.5	20.6	699	2	T05225	extensin homolog F
34	453	20.4	574	2	T43556	Wiskott-Aldrich sy
35	452.5	20.4	378	2	S14959	proline-rich prote
36	452.5	20.4	3534	2	T42567	tegment protein 2
37	452	20.4	574	2	T38819	Wiskott-Aldrich sy
38	446	20.1	461	2	T10741	extensin-like prot
39	445.5	20.1	346	2	S19129	proline-rich prote
40	445	20.1	381	2	S52985	cell wall protein
41	445	20.1	5262	2	T03454	ALR protein - huma
42	443	20.0	429	2	T06296	extensin-like prot
43	441	19.9	393	2	P00479	pistil extensin-11
44	440	19.9	731	2	T04455	hypothetical prote
45	439.5	19.8	426	2	J01696	pistil extensin-11

ALIGNMENTS

RESULT 1
T06291
extensin homolog T9E8.80 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 11-Jun-1999
C:Accession: T06291
R:Bevan, M.; Pohl, T.; Welzenegger, T.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X.; Sch
submitted to the Protein Sequence Database, April 1999
A:Reference number: 215588
A:Accession: T06291
A:Molecule type: DNA
A:Residues: 1-760 <BEV>
A:Cross-references: EMBL:AL049608
A:Experimental source: cultivar Columbia; BAC clone T9E8
C:Genetics:
A:Map position: 4
A:Note: T9E8.80

Query Match 31.0%; Score 687; DB 2; Length 760;
Best Local Similarity 42.1%; Pred. No. 2.2e-25;
Matches 164; Conservative 30; Mismatches 112; Indels 84; Gaps 24;

QY	5	VPPAPAPSPINPVPPPLPA---	ADRLSPVPAPSPDISLAAPLPPDPMP	60		
DB	392	VSPRPV-VNLPPLPSLSPPPADIRTPPTLSPSPSPPV--YSP--	PPPPPP	446		
QY	61	AISALEAPNVPVPPAPGNSAPAPMPPTPLPV--	PPSGAPPVAVAPMPAPK	118		
DB	447	PVYS---PPPPPPPPPVYSPPPPPPPPVYSP-----	PPSPPPPV	494		
QY	119	RMALPPAPAPSPPTSWLAVVPV--	PPVPLPVKMP-----	PSPPVPPPAEPE	170	
DB	495	YSP--PPPPPPPPPV--YSPPPPVYSSPPPPSAPTPVCTRRPPPPPPPPQF	550			
QY	171	TPMPAP-----	PAPLSPSPPPVPPVPPPL--	TLPVVPAPAPANTSNPLRP	223	
DB	551	SPPPPEYVYSSPPPSPPPSPPPSPPPPPYLSP--	PPPPVSSPPPTVYSP	609		
QY	224	APAPPLKGPAP-----	MPAP-----	NSPAP-----	SPSPPVVPTTGGP---	267
DB	610	PPPPCTEPPPPCTEYSPPPPVVYSSPPPVYSSPPPVYSSPPPVVHY	669			
QY	268	-APP-----	NSPP-----	APPAPAPDLGSPAP---	PAPPLPNSAPAPCPAPW	313
DB	670	SSPPPPVHYHSPPSVHYSSPPPPSAPCEBSPPAPRVVHNSPPRVVHNSPP--	PPV	728		
QY	314	PGAPDPAP-----	PLP-----	YSSPPAP	333	
DB	729	HOSPPPSPEYGLPVLVIGVSYASPPPP	758			


```

Db 225 POPTRHAPPTIRHAPPTIQPSLRLLPSPKROQRPPTYSPPPAY---AQSPQDSPT 261
QY 75 PABPGNSAPAPMPPTPLRPVPGSGARPVAVPMPBARKMLPAPAPAPSP 134
Db 282 YSPPTPTSPSPSPISPPPPAYSPSPPTPTPTSPSPPTTSPPTTSPPTT 341
QY 135 TSLAVPVPPVPPVPPVPPVPPVPPVPPVPPVPPVPPVPPVPPVPPV 191
Db 342 SPSISPPSPVSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 399
QY 192 -PPVPPVPPVPPVPPVPPVPPVPPVPPVPPVPPVPPVPPVPPVPPV 241
Db 400 SPPL-PAAP-TYSPPTPTSPSPPTTSPSPPTTSPSPPTTSPSPPTT 457
QY 242 APNSAPAPSPSP---SPPVVF-PTSPSP---PAPEPNSPP-APAPAPLP 290
Db 458 PPAPVAPPPPTTSPSPPTTSPSPPTTSPSPPTTSPSPPTTSPSPPTT 517
QY 291 -----SPAPAP-APPLPNSPAPAPPPAP-PPAP-----DPPAPLP 330
Db 518 RQRPPTPTTSPSPPTTSPSPPTTSPSPPTTSPSPPTTSPSPPTTSPSP 575
QY 331 ----APPACVPAPAPLAPLISGRP 351
Db 576 RQIHSP-PPHROPRTPTTSPSP 598

```

RESULT 5

T31611
 hypothetical protein Y50E8A.g - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
 C:Accession: T31611
 R:Steward, C.
 submitted to the EMBL Data Library, September 1999
 A:Reference number: Z21047
 A:Accession: T31611
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1585 <MIL>
 A:Cross-references: EMBL:AL117200; NID:el549770; PIDN:CAR55050.1; CESP:Y50E8A.g
 C:Experimental source: clone Y50E8A
 C:Genetics:
 A:Gene: CESP:Y50E8A.g
 A:Introns: 25/3; 60/1; 133/2; 217/3; 270/3; 337/2; 400/1; 746/2

Query Match 27.4%; Score 608; DB 2; Length 1585;
 Best Local Similarity 26.2%; Pred. No. 1.6e-21;
 Matches 184; Conservative 15; Mismatches 122; Indels 382; Gaps 25;

```

QY 2 SMVPPAPAPSPSPINPPVPPV-PLPA-----APRTLS 34
Db 913 SAALPPPPPPPP-PPPPAPAPAPSSGSGSAGGSSGSGSAGGSSGSA 970
QY 35 PVPAPAPSPISLAAPLPDPMPAPAWS-----ALEAP 69
Db 971 PPPPPPPPP-PPPPAPAPAPAPSSGSGSAGGSSGSGSAGGSSGSA 1026
QY 70 NPVPAPAPPGNSAPAPM-----PPTPPLPPVPP 99
Db 1027 PPPPPPPPPAPAPAPSSGSGSAGGSSGSGSAGGSSGSGSTAPPPPPPP 1086
QY 100 GSGAPRPV-----AVPMPAPAKRMALPAPAPAP 130
Db 1087 PAAPAPAPAPSSGSGSAGGSSGSGSAGGSSGSGSAGGSSGSA 1140
QY 131 PSPPTG-----WLAVPVPPVPPVPPVPPVPPVPPVPPVPPVPPVPP 158
Db 1141 APAPSSGSGSAGGSSGSGSAGGSSGSGSAGGSSGSGSAGGSSG 1200
QY 159 -----PPVPPPPAPPETNPPAPAPPLENS----- 185

```

```

Db 1201 SGGSSGSGSAGGSGSGSAGGSSAGGSSAGGSSAGGSSAGGSSAGG 1255
QY 186 -----PPPPVPPVPPVPPVPPVPPVPPVPPVPPVPPVPPVPPV 215
Db 1256 SGGSAGGSGSGGYTGSAAPPPPPPPPP-----PPAPAPAPAPSSG 1308
QY 216 -----SNSPLRPPAPAPPLKPGPPAPPMPPAPANS----- 245
Db 1309 GGSAGGSGSGSGGYSGGSAAPPPPPPP-----PPAPAPAPAPSSG 1366
QY 246 -----PAAPSPSPVVPPTPPGPPAP-PPAPSS----- 275
Db 1367 GGGSSGSGYTGSAPPPPPPP-----PPPPAPAPAPSSGSGSGS 1422
QY 276 -----PPAPAPAPAPLP-----GPPSPA 294
Db 1423 SGGYSGGSAAPPPPPAPAPAPAPSSGSGSAGGSSGSGSAGGSSG 1482
QY 295 PPAPPLPNSPAPPPGPA-----WPGAPDPAPAPLPYSSP 329
Db 1483 PPPPPPPAPAPAPAPSSGSGSAGGSSGSGSAGGSSGSGYTGSAP 1537
QY 330 PAPACVPAPAPLAPLISGRPSNSWGVFTMLSRPSNGAAA 372
Db 1538 PPPPPAPAP-----APAPSSGSGSG-----SSGSA 1565

```

RESULT 6

T04859
 extensin homolog F28A21.80 - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse ear cress)
 C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-Jun-1999
 C:Accession: T04859
 R:Bevan, M.; Mueller, M.W.; Mendenhall, A.; Felber, R.; Bancroft, I.; Mewes, H.W.; May
 submitted to the Protein Sequence Database, February 1999
 A:Reference number: Z15387
 A:Accession: T04859
 A:Molecule type: DNA
 A:Residues: 1-839 <BBV>
 A:Cross-references: EMBL:AL035526
 A:Experimental source: cultivar Columbia; BAC clone F28A21
 C:Genetics:
 A:Map position: 4
 A:Introns: 623/3
 A:Note: F28A21.80

Query Match 25.8%; Score 571.5; DB 2; Length 839;
 Best Local Similarity 30.7%; Pred. No. 4.6e-20;
 Matches 150; Conservative 36; Mismatches 127; Indels 175; Gaps 22;

```

QY 1 NSMVPPAPAPSP-----PINPPVPPVPPVPPVPPVPPVPPVPPVPP 37
Db 370 NCLPAPAPAPSSGSGSAGGSSGSGSAGGSSGSGSAGGSSGSA 425
QY 38 PPAPPS-PISLAAPLPDPMPAPAWSALEAPNPVPPAPPGNSAPAPMPPT 96
Db 426 PSISPPSPITVPPSPPTTSPSGSP-----PSPSIVSP-PPSTPSPSP 475
QY 97 VPPGSGAPRPVPPAP-PPAPAPAKRMALPAPAPSPSPISLAAPVPPVPP 155
Db 476 TPGGSPSSPTTPTTTPGGSPSPSPPTTTPGGSPSPSPPTTTPGGSP 535
QY 156 PPSPPVPPPPAPPETNPP-----PAPAP-----PLNSPP-----PP 196
Db 536 PSTPTPTGSPSPSSPTSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 595
QY 197 VPPLTNPVPPAPAPAP-TPNSPLRPPAPAP----- 228
Db 596 SPPL-PPVIPSPPVIGPPSPSP-PPSTPTPTGTLNPPHLLPQLSHLP 650
QY 229 -----PL-----KPPPPAP----- 237

```

Db 651 HILPSRHHLLRRKHTIHRNHLHNDRNLQETALRLPLLYLCRHRLLASPPAPAYYY 710
Qy 238 --PMPAPNSAPAP-----SPSPVYVFPPR-----GPPAPEPSSSPA 278
Db 711 SSPQRPPIHYSLPRTTYTHYISPPPTPIHSPPOSHPCICXSPPTVYVYNNPP 770
Qy 279 PPAAP-AAALPGP-----SPAPPA-----PPLPNSPAAPGPPAPGAPDPAPPLP 325
Db 771 PPSAHYSPSPSPRYYYNSPPRPRAVHYSPSPRPVYHNSQRPPIYEG-PLPPIPGIS 829
Qy 326 YSSPPAP 333
Db 830 YASPPPP 837

RESULT 7
T10863
extensin precursor - kidney bean
Alternate names: cell wall protein; hydroxyproline-rich glycoprotein
Species: Phaseolus vulgaris (Kidney bean)
Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000
C:Accession: T10863
R:Wycoff, K.L.; Powell, P.A.; Gonzales, R.A.; Corbin, D.R.; Lamb, C.; Dixon, R.A.
Plant Physiol. 109, 41-52, 1995
A:Title: Stress activation of a bean hydroxyproline-rich glycoprotein promoter is super-
A:Reference number: 217192; MUID:96061709
A:Accession: T10863
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-580 <WYC>
A:Cross-references: EMBL:U18791; NID:9272263; PIDN:AAA87902.1; PID:9272264
C:Superfamily: hydroxyproline-rich glycoprotein
C:Keywords: glycoprotein; hydroxyproline
F1:29/Domain: signal sequence #status predicted <SIG>
F1:30-580/Product: extensin #status predicted <MAT>

Query Match 25.2%; Score 571; DB 2; Length 580;
Best Local Similarity 33.8%; Pred. No. 3.7e-20;
Matches 166; Conservative 29; Mismatches 152; Indels 144; Gaps 30;

Qy 1 NSMPVP--PAPR--APSPINPVPPV-----PP--LPAPRTLSPPVP--APPS 43
Db 61 HSPRPKHSPPRYYYHNSPPRPKHSPPRYYYHNSPPRPKHSPPRYYYHNSPPRPKHSPP 120
44 PISLAAPRLPDPRRAIWSALEAP-NRPVR--ARPGNSAPAPR--MPTPRL 94
121 PYYHNSPPRPKHSPPRYYYHNSPPRPKHSPPRYYYHNSPPRPKHSPPRYYYHNSPP 180
Qy 95 PVPVP--GSGAPRPVAPVP-----MPRAPKMPA-----LPAPRAPSPRTSLVLP 141
Db 181 SPSPRYHNSPPRPKHSPPRYYYHNSPPRPKHSPPRYYYHNSPPRPKHSPPRYYYHNSP 240
Qy 142 VPPVPVRLVUKMPPSP--PVPFP-----PRAEP-----ETPNP-----P 175
Db 241 PRKHSPPRYYYHNSPPRPKHSPPRYYYHNSPPRPKHSPPRYYYHNSPPRPKHSPP 300
Qy 176 APAPRLNSPPRPV-----PPV-----PVPRLTLNPPVPA-----PRAANTSN 217
Db 301 SPSPRYHNSPPRPKHSPPRYYYHNSPPRPKHSPPRYYYHNSPPRPKHSPPRYYYHNSP 360
Qy 218 SPLRP--PAPAPRLK-RGP-----PAPRW-----PRAANSAPAPSP--PSP 255
Db 361 SPSPRYHNSPPRPKHSPPRYYYHNSPPRPKHSPPRYYYHNSPPRPKHSPPRYYYHNSP 420
Qy 256 PVPVP--PVP-----GPPAPEPNSPPAP--PAPAPALRGSPAP--PAPAPALRGSPAP 295
Db 421 PVPYKYSPSPRYYYHNSPPRPKHSPPRYYYHNSPPRPKHSPPRYYYHNSPPRPKHSPP 480
Qy 296 -----PAPPLPNSPAPDPGPPAPMGADPP-----ADPLP--YSSPPAPACVPGAP 341

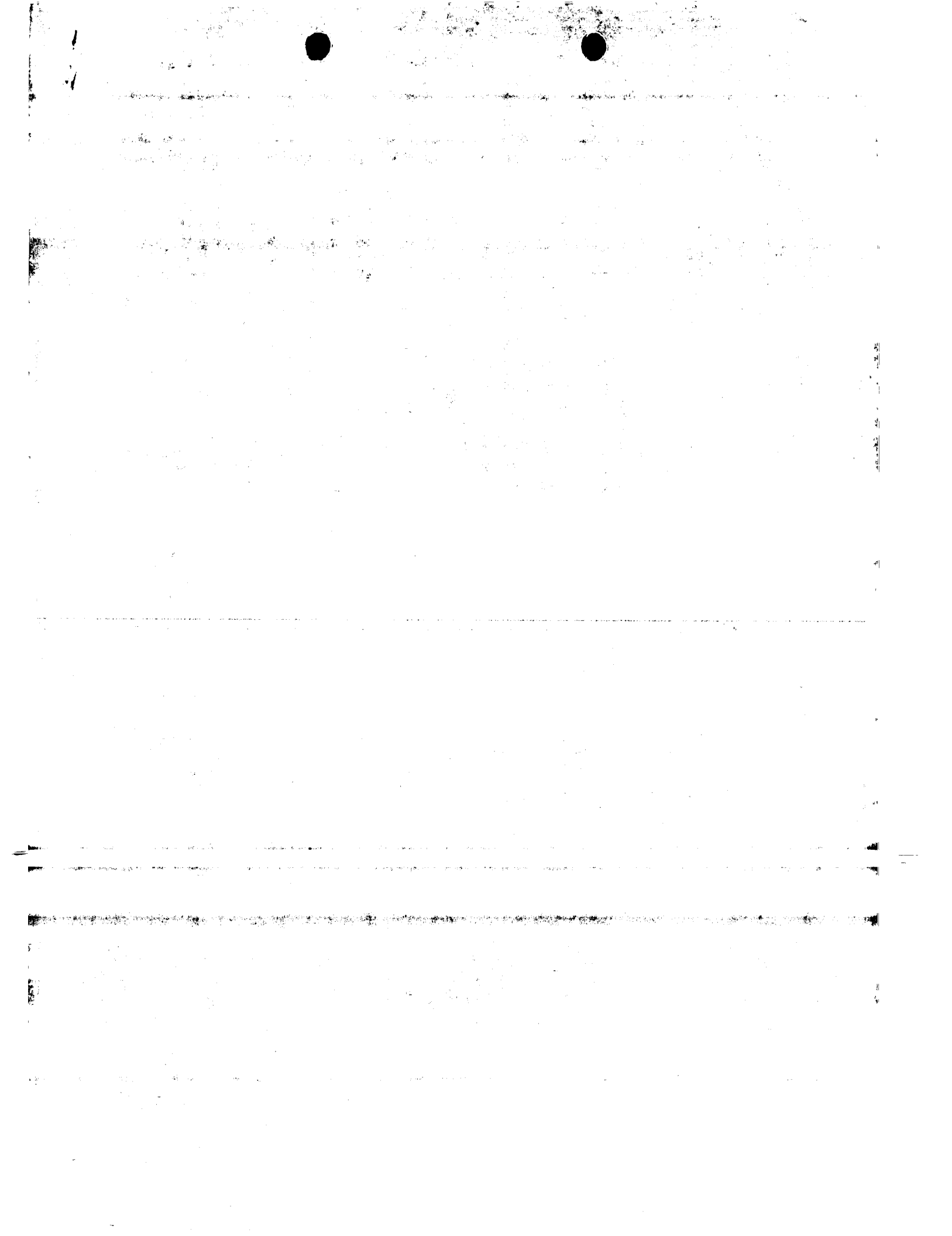
Db 481 PVPYKYSPSPRYYYHNSPPRPKHSPPRYYYHNSPPRPKHSPPRYYYHNSPPRPKHSPP 540
Qy 342 L-----APLPI 347
Db 541 YKYKSPSPPV 551

RESULT 8
S25298
extensin (clone Tom J-10) - tomato
C:Species: Lycopersicon esculentum (tomato)
C:Date: 03-Feb-1994 #sequence_revision 03-Feb-1994 #text_change 21-Jul-2000
C:Accession: S25298
R:Zhou, J.; Rumeau, D.; Showalter, A.M.
Plant Mol. Biol. 20, 5-17, 1992
A:Title: Isolation and characterization of two wound-regulated tomato extensin genes.
A:Reference number: S25298; MUID:92385769
A:Accession: S25298
A:Molecule type: DNA
A:Residues: 1-388 <ZHO>
A:Cross-references: EMBL:M76670; NID:9170441; PIDN:AAA34163.1; PID:9170442
C:Superfamily: hydroxyproline-rich glycoprotein
C:Keywords: cell wall; glycoprotein; hydroxyproline

Query Match 25.2%; Score 558; DB 2; Length 388;
Best Local Similarity 36.5%; Pred. No. 1.1e-19;
Matches 149; Conservative 27; Mismatches 140; Indels 92; Gaps 23;

Qy 4 PVP-----PAPAPSPINPVPPV-----PLPAPRTL--SPVPAPPSPISLAAP 50
Db 7 PTPYKSPSPSPSPRYYYHNSPPRPKHSPPRYYYHNSPPRPKHSPPRYYYHNSPPRPKHSPP 62
Qy 51 PVPVP-----PMPAPISALEAPNPPV-----PAPPGNSAPAPMPTPRL--P 95
Db 63 PVPSPRYYYHNSPPRPKHSPPRYYYHNSPPRPKHSPPRYYYHNSPPRPKHSPPRYYYHNSP 122
Qy 96 PVPSSGAPRPVAPVMPAPKMPALPAPAPSPRTSLAVPVPVPPVPLPVKM 155
Db 123 PVP--SPKRYKYSPP--PSPKRYKYSPP--PSPKRYKYSPP--PSPKRYKYSPP 170
Qy 156 PPSPPVPFPPEPPTPMP-----PAPAPRLNSPPRPV-----PVPVPPRLTLN 203
Db 171 PP-PPSPRYYYHNSPPRPKHSPPRYYYHNSPPRPKHSPPRYYYHNSPPRPKHSPPRYYYK 229
Qy 204 PVPAPAPAPNTSNPLPAPAPAPRLKGPAPAPMPAPAPAPSPSPVPV--P 260
Db 220 SPSPSPRYYYKSP-----PSPRYYYKSPSPSPSPRYYYKSPSPSPSPRYYYKSP 284
Qy 261 PTPPGR-SAPPEPNSPPAP--PAP-----AAPLGGSPAP--PAPPLPNSPAP 307
Db 285 PVPSPSPRYYYHNSPPRPKHSPPRYYYHNSPPRPKHSPPRYYYHNSPPRPKHSPPRYYYHNSP 341
Qy 308 PGPAPMGADPPA--PPLYSPPAPAPACVPGADLAPLPIISGR 351
Db 342 --PVPYHNSPPRPVNSPPRPVYSSPPRPVKSPP--PVIYIGSP 382

RESULT 9
T06782
extensin - soybean
N:Alternate names: hydroxyproline-rich glycoprotein
C:Species: Glycine max (soybean)
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-May-2000
C:Accession: T06782
R:Ahm, J.H.; Choi, Y.; Kwon, Y.M.; Kim, S.G.; Choi, Y.D.; Lee, J.S.
Plant Cell 8, 1477-1490, 1996
A:Title: A novel extensin gene encoding a hydroxyproline-rich glycoprotein requires s
A:Reference number: 215809; MUID:96434536
A:Accession: T06782
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-432 <AHN>



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 5, 2001, 14:27:36 ; Search time 45.19 seconds
(without alignments)
153.384 Million cell updates/sec

Title: US-09-461-774-10

Percent score: 2216

Sequence: 1 NSMPVPPAPAPPPSPINPPV.....NGAAMASALAVAPAPKVV 386

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 174772 seqs, 17957048 residues

Minimum number of hits satisfying chosen parameters: 174772

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents.AA:*
1: /cgn2_6/prodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/prodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/prodata/2/1aa/6.COMB.pep:*
4: /cgn2_6/prodata/2/1aa/PTCUTS.COMB.pep:*
5: /cgn2_6/prodata/2/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	832	37.5	408	1	US-07-609-716-65 Sequence 65, Appl
2	832	37.5	408	3	US-08-475-411A-65 Sequence 65, Appl
3	721	32.5	330	1	US-08-642-255-32 Sequence 32, Appl
4	598.5	27.0	504	3	US-09-219-849-3 Sequence 3, Appl
5	598.5	27.0	561	1	US-08-642-255-52 Sequence 52, Appl
6	570.5	25.7	720	3	US-09-219-849-4 Sequence 4, Appl
7	570.5	25.7	777	1	US-08-642-255-53 Sequence 53, Appl
8	560	25.3	357	1	US-07-609-716-66 Sequence 66, Appl
9	560	25.3	357	1	US-08-642-255-33 Sequence 33, Appl
10	560	25.3	357	3	US-08-475-411A-66 Sequence 66, Appl
11	540.5	24.4	960	3	US-09-219-849-5 Sequence 5, Appl
12	528.5	23.8	1064	1	US-08-642-255-62 Sequence 62, Appl
13	479.5	21.6	829	1	US-08-642-255-132 Sequence 132, Appl
14	479.5	21.6	829	1	US-08-397-633A-53 Sequence 53, Appl
15	478	21.6	837	1	US-08-175-155-68 Sequence 68, Appl
16	478	21.6	837	1	US-08-477-509B-103 Sequence 103, Appl
17	478	21.6	837	1	US-08-642-255-101 Sequence 101, Appl
18	478	21.6	837	2	US-08-707-237A-75 Sequence 75, Appl
19	478	21.6	837	3	US-08-482-085B-103 Sequence 103, Appl
20	478	21.6	897	1	US-08-397-633A-50 Sequence 50, Appl
21	472	21.3	417	1	US-08-175-155-69 Sequence 69, Appl
22	472	21.3	417	1	US-08-477-509B-104 Sequence 104, Appl
23	472	21.3	417	1	US-08-642-255-102 Sequence 102, Appl
24	472	21.3	417	2	US-08-707-237A-76 Sequence 76, Appl
25	472	21.3	417	3	US-08-482-085B-104 Sequence 104, Appl
26	452	20.4	234	1	US-08-642-255-51 Sequence 51, Appl
27	431.5	19.5	682	1	US-08-642-255-126 Sequence 126, Appl
28	431.5	19.5	682	1	US-08-397-633A-36 Sequence 36, Appl

29	424.5	19.2	297	2	US-08-580-545B-6 Sequence 6, Appl
30	424.5	19.2	297	3	US-09-262-653A-6 Sequence 6, Appl
31	411.5	18.6	334	5	5202236-3 Patent No. 5202236
32	410.5	18.5	331	5	5202236-37 Patent No. 5202236
33	403.5	18.2	761	2	US-08-707-237A-84 Sequence 84, Appl
34	403.5	18.2	762	1	US-08-642-255-114 Sequence 114, Appl
35	403.5	18.2	762	1	US-08-642-255-120 Sequence 120, Appl
36	403.5	18.2	762	1	US-08-397-633A-26 Sequence 26, Appl
37	403.5	18.2	762	1	US-08-397-633A-31 Sequence 31, Appl
38	402.5	18.2	1255	2	US-09-080-897-4 Sequence 4, Appl
39	402.5	18.2	1255	3	US-08-899-595-1 Sequence 1, Appl
40	390	17.6	1315	3	US-08-899-595-3 Sequence 3, Appl
41	388	17.5	1248	2	US-09-080-897-2 Sequence 2, Appl
42	383.5	17.3	252	1	US-08-642-255-61 Sequence 61, Appl
43	376	17.0	306	1	US-08-217-327-6 Sequence 6, Appl
44	373	16.8	960	3	US-09-219-849-6 Sequence 6, Appl
45	372.5	16.8	826	1	US-07-638-431-2 Sequence 2, Appl

ALIGNMENTS

```
RESULT 1
US-07-609-716-65
; Sequence 65, Application US/07609716
; Patent No. 5314581
;
; GENERAL INFORMATION:
; APPLICANT: Ferrari, Franco A.
; APPLICANT: Cappello, Joseph
; TITLE OF INVENTION: Functional Recombinantly Prepared
; TITLE OF INVENTION: Synthetic Protein Polymer
; NUMBER OF SEQUENCES: 118
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Holbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: US
; ZIP: 94111
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/609,716
; FILING DATE: 06-NOV-1990
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Rowland, Berttram I
; REGISTRATION NUMBER: 20015
; REFERENCE/DOCKET NUMBER: A-53186-3/BIR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249
;
; INFORMATION FOR SEQ ID NO: 65:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 408 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-609-716-65
;
; Query Match 37.5%; Score 832; DB 1; Length 408;
; Best Local Similarity 46.8%; Pred No. 7.6e-46;
; Matches 169; Conservative 5; Mismatches 155; Indels 32; Gaps 11;
;
; QY 4 PVPAPAPPPSPINPPVPLPA----APRTLSPPVPPAPPSPTSLAAPLPDPMP 59
; Db 42 PGPPGPPGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 101
```

```

QY 60 PAWSAL---EAPNPVPVPPAGNSAPAPMPPTPLPPVPPGS-GAP-RFVAVPMP 114
Db 102 PGPAPVSPAPGPPGPPGPGAGVGPAGPAGVSGAPAGPAGPAGP 161
QY 115 PAKKMPALPPAPAPSPPTSWLAVPVVPPVPLVKMPSPPVPPFAEETPNP 174
Db 162 PGPAPAG-PGPGPPGPP-----GPAAGVSGAPAGPAGPAGPAGPAGP 212
QY 175 PAPPAPLENP-----PPPPVPPVPPVPLTLNPPVPPAPPAANTNSPLRPAPAP 228
Db 213 PGPAPGPPAGPVSGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 272
QY 229 PLKGPAPAPMPAPNSAPAPSPSPVPPVPPPPGPPAPPEPNSPPAPAPAPLP 288
Db 273 PGPAPGPPGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 326
QY 289 GSPAPAPAPLPNSAPAPGPPAPAGAPDPAPPLPYSSPPAPAPACVPAPLAPLPIS 348
Db 327 -PGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 384
349 G 349
Db 385 G 385

RESULT 2
US-08-475-411A-65
Sequence 65, Application US/08475411A
Patent No. 6140072
GENERAL INFORMATION:
APPLICANT: Ferrari, Franco A.
TITLE OF INVENTION: Functional Recombinantly Prepared
NUMBER OF SEQUENCES: 119
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: CA
COUNTRY: US
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/475,411A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION NUMBER:
APPLICATION NUMBER: US 07/609,716
FILING DATE: 06-NOV-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/269,429
FILING DATE: 09-NOV-1988
APPLICATION NUMBER: US 07/114,618
FILING DATE: 29-OCT-1987
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 06/927,258
FILING DATE: 04-NOV-1986
ATTORNEY/AGENT INFORMATION:
NAME: Trecartin, Richard F.
REGISTRATION NUMBER: 31,801
REFERENCE/DOCKET NUMBER: A-55186-9/RFT/MTK
TELEPHONE: 415-781-1989
TELEFAX: 415-398-3249
INFORMATION FOR SEQ ID NO: 65:
SEQUENCE CHARACTERISTICS:
LENGTH: 408 amino acids

```

```

TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-475-411A-65

Query Match 37.5
Best Local Similarity 46.3
Matches 169; Conservative

QY 4 PVPAPAPAPSPINPPVPP 59
Db 42 PGPAPGPPAGPPGPPGPP 101
QY 60 PAWSAL---EAPNPVPV 114
Db 102 PGPAPVSGAPAGPAGPAG 161
QY 115 PAKKMPALPPAPAPSPPT 174
Db 162 PGPAPAG-PGPGPPGPP 212
QY 175 PAPPAPLENP-----PI 228
Db 213 PGPAPGPPAGPVSGAPAG 272
QY 229 PLKGPAPAPMPAPNSPA 288
Db 273 PGPAPGPPGAPGPPGPP 326
QY 289 GSPAPAPAPLPNSAPAP 348
Db 327 -PGPPGPPGPPGPPGPP 384
349 G 349
Db 385 G 385

RESULT 3
US-08-642-255-32
Sequence 32, Application US,
Patent No. 5773249
GENERAL INFORMATION:
APPLICANT: FERRARI, Fra
TITLE OF INVENTION: Pro
NUMBER OF SEQUENCES: 131
CORRESPONDENCE ADDRESS:
ADDRESSEE: FLEHR, HOH
STREET: 4 Embarcadero
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy d
OPERATING SYSTEM: PC-
SOFTWARE: Patentin Re
CURRENT APPLICATION DATA
APPLICATION NUMBER: U
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATI
NAME: ROWLAND, Berta
REGISTRATION NUMBER:
REFERENCE/DOCKET NUMBE
TELECOMMUNICATION INFORM
TELEPHONE: (415) 494-87
TELEFAX: (415) 494-87

Score 832; DB 3; Length 408;
Pred. No. 7.6e-46;
5; Mismatches 155; Indels 32; Gaps 11;

```

TELEX: 910 277299 FHT UR
 INFORMATION FOR SEQ ID NO: 32:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 330 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-642-255-32

Query Match 32.5%; Score 721; DB 1; Length 330;
 Best Local Similarity 45.3%; Pred. No. 5, 8e-39;
 Matches 150; Conservative 5; Mismatches 120; Indels 56; Gaps 12;

QY 9 PPAAPSPINPPVPPVPLPAAPRTLSPPVPAPPPSPISLAAPLPDPMPAPPAIWSALEA 68
 DB 26 PPAASDMGAPGPPGPP-----GPPGPPGAPGP---GPPGPPGPPGPPGPPGPPGPPG 74
 QY 69 PNPVPAPPPGPPSAPAPMPPTPLPPVPPGSGAPPPVAVPMPAPKRMALPPAP 128
 DB 75 PGPDPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 120
 QY 129 APPSPPTSWLAVVPPVPPVPLVKMPSPVPPPPPPAPETPPNPAPPPAPPLENSPP 188
 DB 121 GPPGPP-----GAPGPPGPPGPP-----GPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 172
 QY 189 PVPVPPVPPVPLTLNPPVPPAPPAANTSNSPLRPAPADPDLKGPAPPPAPPPAPNSPAA 248
 DB 173 PPG 223
 QY 249 P-----PSPSPVPPVPPPPGPPAPPEPNSSPPAPAP--PAAPLPGSPSPAPPPPL 300
 DB 224 PVGSPGAPGPPGPPGPP--PGPPGAPGPP-----GPPGPPGPPGPPGPPGPPGPPGPPG 277
 QY 301 PNSPAPPPGPPAMPGAPDPAPPLPYSSPPA 331
 DB 278 PPGPPGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 308

RESULT 4
 US-09-219-849-3
 ; Sequence 3, Application US/09219849
 ; Patent No. 6150081
 ; GENERAL INFORMATION:
 ; APPLICANT: VAN HEERDE, GEORGE V.
 ; APPLICANT: VAN RIJN, ALEXIS C.
 ; APPLICANT: BOUMSTRA, JAN B.
 ; APPLICANT: DE WOLF, FREDERIK A.
 ; APPLICANT: MOOBROEK, ANDREAS
 ; APPLICANT: WERTEN, MARC W.T.
 ; APPLICANT: WIND, RICHELIE D.
 ; APPLICANT: VAN DEN BOSCH, TANJA J.
 ; TITLE OF INVENTION: SILVER HALIDE EMULSIONS WITH RECOMBINANT COLLAGEN
 ; TITLE OF INVENTION: SUITABLE FOR PHOTOGRAPHIC APPLICATION AND ALSO THE
 ; TITLE OF INVENTION: PREPARATION THEREOF
 ; FILE REFERENCE: 2728-2
 ; CURRENT APPLICATION NUMBER: US/09/219,849
 ; CURRENT FILING DATE: 1998-12-23
 ; NUMBER OF SEQ ID NOS: 50
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 3
 ; LENGTH: 504
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence: Illustrative
 ; OTHER INFORMATION: amino acid sequence
 US-09-219-849-3

Query Match 27.0%; Score 598.5; DB 3; Length 504;
 Best Local Similarity 35.6%; Pred. No. 4e-31;

Matches 154; Conservative 3; Mismatches 200; Indels 75; Gaps 14;
 QY 4 PVPAPAPPPSPINPPVPPVPP--LPAAPRTLSPPVPAPPPSPISLAAPLPDPMPAPPAI 62
 DB 21 PGPAGPPGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 80
 QY 63 WSALEAPNPVPPAPPGPPNSAPAPVMPPTPLPPVPPGSGAP-----RPVAVPMPMPAP 117
 DB 81 PGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 140
 QY 118 KRM-----PALPPAPPPAPSPPTSWLAVVPPVPPVPLP-----VKMPSPVPPPPPA 167
 DB 141 AGPPGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 196
 QY 168 EPETPNPAPPPAPPLENSPPPPPPVPPVPPVPLTLNPP-----V 206
 DB 197 PPGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 256
 QY 207 PPAAPPAANTSNSPLRPAPPPAPPLKGPAPPPAPMPAPNSPAPPPSPPP-----VPV 259
 DB 257 PGPCKAHGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 316
 QY 260 FPPPPGPPAPPEPNSSP--PAPPPAPAPLP-----GSPPPAPPPAPPLNSPAPPPGPPA 312
 DB 317 APGPAGPPGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 376
 QY 313 WPGAPDPAPPLPYSSPPAPPPACVP--GAPLAPPLISGPPSNVGVFTMLSRPSSGAA 371
 DB 377 PPGAPGPPGPP--GAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 418
 QY 372 AASALAVAPAPA 383
 DB 419 PGPPPG--APGPA 429

RESULT 5
 US-08-642-255-52
 ; Sequence 52, Application US/08642255
 ; Patent No. 5773249
 ; GENERAL INFORMATION:
 ; APPLICANT: CAPPELLO, Joseph
 ; APPLICANT: FERRARI, Franco A.
 ; TITLE OF INVENTION: High Molecular Weight Collagen-Like
 ; TITLE OF INVENTION: Protein Polymers
 ; NUMBER OF SEQUENCES: 135
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT
 ; STREET: 4 Embarcadero Center, Suite 3400
 ; CITY: San Francisco
 ; STATE: California
 ; COUNTRY: USA
 ; ZIP: 94111-4187
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/642,255
 ; FILING DATE:
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: ROWLAND, Bertlam I.
 ; REGISTRATION NUMBER: 20,015
 ; REFERENCE/DOCKET NUMBER: A55556-3/BIR
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (415) 494-8700
 ; TELEFAX: (415) 494-8771
 ; TELEX: 910 277299 FHT UR
 ; INFORMATION FOR SEQ ID NO: 52:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 561 amino acids
 ; TYPE: amino acid

STRANDEDNESS: Single
TOPOLOGY: Linear
MOLECULE TYPE: Protein
US-08-642-255-52

Query Match 27.0%, Score 598.5; DB 1; Length 561;
Best Local Similarity 35.6%, Pred. No. 4,4e-31;
Matches 154; Conservative 3; Mismatches 200; Indels 75; Gaps 14;

QY 4 PVPPAPAPSPINPPVPP-LEPAARTLSPPVPPAPSPISLAAPPLPDDPPMPAI 62
DB 54 PGPAPGPAPGPAPGPAPGPAPGPAPGPAPGPAPGPAPGPAPGPAPGPAPGP 113
QY 63 WSALEAPNPVPVPPAPGNSAPAPMPTPLPVPVPGSGAP-----RPVAVPPMPAP 117
DB 114 PGAPGPAPGPAPGPAPGPAPGPAPGPAPGPAPGPAPGPAPGPAPGPAPGP 173
QY 118 KRM-----PALPPAPAPAPSPPTSLAVPPVPPVPPPL-----YKMPSPVPPPPA 167
DB 174 AGPPAPGPAPGPAPGPAPGP-----GAPGAPGPAPGPAPGPAPGPAPGP 229
QY 168 EPPTNPAPAPAPPLPPPPPPPPPPPPPLTINPP-----V 206
DB 230 PPGAPGPAPGPAPGPAPGPAPGPAPGPAPGPAPGPAPGPAPGPAPGPAPG 289
QY 207 PPAPPAFTNSPLPPAPAPPLKGPAPPPAPPPAPNSPAPSPSP-----VPV 259
DB 290 PGPAPGPAPGPAPGPAPGPAPGPAPGPAPGPAPGPAPGPAPGPAPGPAPG 349
QY 260 FPPTPPAPAPPPNSP--PAPAPAPAPLP-----GSPAPAPAPPLPNSPAPPPA 312
DB 350 APGPAPGPAPGPAPGPAPGPAPGPAPGPAPGPAPGPAPGPAPGPAPGP 409
QY 313 WPGAPDPPAPPLPYSSPPAPPPACVP--GAPLAPPLISGRPSNSWGVFTMLSRPSNGAA 371
DB 410 PPGAPGPAPGP--GAPGPAPGPAPGPAPGPAPGP--AGPP-----GAPG 451
QY 372 AASALAVAPAPA 383
DB 452 PGPAPGP-APGPA 462

RESULT 6
US-09-219-849-4
Sequence 4, Application US/09219849
Patent No. 6150081

GENERAL INFORMATION:

APPLICANT: VAN HERDE, GEORGE V.
APPLICANT: VAN RIJN, ALEXIS C.
APPLICANT: BOOMSTRA, JAN B.
APPLICANT: DE WOLF, FREDERIK A.
APPLICANT: MOOBROEK, ANDREAS
APPLICANT: WERTEN, MARC W.T.
APPLICANT: WIND, RICHEL D.
APPLICANT: VAN DEN BOSCH, TANJA J.
TITLE OF INVENTION: SILVER HALIDE EMULSIONS WITH RECOMBINANT COLLAGEN
TITLE OF INVENTION: SUITABLE FOR PHOTOGRAPHIC APPLICATION AND ALSO THE
FILE REFERENCE: 2728-2
CURRENT APPLICATION NUMBER: US/09/219,849
CURRENT FILING DATE: 1998-12-23
NUMBER OF SEQ ID NOS: 50
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 4
LENGTH: 720
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Illustrative
US-09-219-849-4

Query Match 25.
Best Local Similarity 34.
Matches 145; Conservative 6; Mismatches 209; Indels 67; Gaps 11;

QY 4 PVPPAPAPSPINPPVPP-LEPAARTLSPPVPPAPSPISLAAPPLPDDPPMPAI 62
DB 30 PGPAPGPAPGPAPGPAPGPAPGPAPGPAPGPAPGPAPGPAPGPAPGPAPG 89
QY 63 WSALEAPNPVPVPPAPGNSAPAPMPTPLPVPVPGSGAP-----RPVAVPP 112
DB 90 PGPAPGPAPGPAPGPAPGPAPGPAPGPAPGPAPGPAPGPAPGPAPGPAPG 149
QY 113 MPAPAPAPPLPPAPAPAP-----GAPGAPGPAPGPAPGPAPGPAPGP 201
DB 150 AGPKAHGAPGPAPGPAPGPAPGPAPGPAPGPAPGPAPGPAPGPAPGPAPGP 232
QY 173 NPAPAPAPPLPP 261
DB 202 GPAPGPAPGPAPGPAPGPAPGPAPGPAPGPAPGPAPGPAPGPAPGPAPG 262
QY 233 GPPAPMPAPNSPAP--PAP-----PSPPSPVPPVPPPT 262
DB 262 GAPGPAPGPAPGPAPGPAPGPAPGPAPGPAPGPAPGPAPGPAPGPAPGP 319
QY 263 PGPAPAPPPNSP--PAP-----GSPAPAPAPPLPNSPAPPPAPMPG 315
DB 320 PGPAPGPAPGPAPGPAPGPAPGPAPGPAPGPAPGPAPGPAPGPAPGPAPG 379
QY 316 APDPAPAPPLPYSSPPAPPA-----GAPLAPPLISGRPSNSWGVFTMLSRPSNGAAA 374
DB 380 APGPAPGP--GAPGPAPGPAPGPAPGPAPGPAPGPAPGPAPGPAPGPAPGP 431
QY 375 ALAYAPA 381
DB 432 KGAGHPA 438

RESULT 7
US-08-642-255-53
Sequence 53, Application US,
Patent No. 5773249

GENERAL INFORMATION:

APPLICANT: CABELLO, JO
APPLICANT: FERRARI, FRA
TITLE OF INVENTION: Hig
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: FLEHR, HOH
STREET: 4 Embarcadero
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy d
COMPUTER: IBM PC comp
OPERATING SYSTEM: PC-
SOFTWARE: Patentin Re
CURRENT APPLICATION DATA
APPLICATION NUMBER: U
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: ROWLAND, BETTE
REGISTRATION NUMBER:
REFERENCE/DOCKET NUMB
TELECOMMUNICATION INFORM
TELEPHONE: (415) 494-87
TELEFAX: (415) 494-87
TELEX: 910 277299 FHT
INFORMATION FOR SEQ ID NO:

h
A.
Molecular weight collagen-like
n Polymers

H. TEST, ALBRITTON & HERBERT
inter, Suite 3400

ble
/MS-DOS
se #1.0, Version #1.30
8/642,255

015
A55556-3/BIR

ON:

13:

SEQUENCE CHARACTERISTICS:
LENGTH: 777 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-642-255-53

Query Match 25.7%; Score 570.5; DB 1; Length 777;
Best Local Similarity 34.0%; Pred. No. 3.3e-29;
Matches 145; Conservative 6; Mismatches 209; Indels 67; Gaps 11;

QY 4 PVPPAPPAPSPINPVPPVPP--LPAPRTLSPPVPPAPSPISLAAPPLPPDPMPRAI 62
DB 63 PGAPGPPGAPGPPGAPGPPGAPGPPGAPGPPGAPGPPGAPGPPGAPGPPGAP 122
QY 63 WSALEAPNPVPPAPPGPNSAPAPMPPTPLPPVPPSGAPR-----PVPAVPP 112
DB 123 PGAPGPPGAPGPPGAPGPPGAPGPPGAPGPPGAPGPPGAPGPPGAPGPPGAP 182
QY 113 MPAPKMPPLPPAPAPSPSPISLAVPVPPVPPVPLPVKMPSPSPVPPPPAPETP 172
DB 183 AGKGAHGPAPGPPGAPGPPGAPGPPGAPGPPGAPGPPGAPGPPGAPGPPGAP 234
QY 173 NPAPAPAPLENSPPPPVPPVPPVPLTLNPVPPAPAPANTSNSPLRPPAPAPPLKP 232
DB 235 GPAPGPPGAPGPPGAPGPPGAPGPPGAPGPPGAPGPPGAPGPPGAPGPPGAP 294
QY 233 GPAPAPPPAPNSPAP-----PSPSPSPVPPVPT 262
DB 295 GAGPPAPPPGAPGPPGAPGPPGAPGPPGAPGPPGAPGPPGAPGPPGAPGPPGAP 352
QY 263 PGCPAPDEPNSSD--PAPAPAPAPLP-----GPSPAPDPAPPLPNSPAPPGPPAPMG 315
DB 353 PACPPGAPGPPGAPGPPGAPGPPGAPGPPGAPGPPGAPGPPGAPGPPGAPGPPG 412
QY 316 APPPPAPPLYSPPAPAPCPVP--GAPLAPPLISGRPSNSWCVFTLSRPSNKAANAAS 374
DB 413 APGAPPP--GAPGPPGPPGAPGPPGAPGPPGAPGPPGAPGPPGAPGPPGAPGPP 464
QY 375 ALAYAPA 381
DB 465 KGAHGPA 471

RESULT 8
US-07-609-716-66
Sequence 66, Application US/07609716
Patent No. 5514581

GENERAL INFORMATION:
APPLICANT: Ferrari, Franco A.
APPLICANT: Cappello, Joseph
TITLE OF INVENTION: Functionally Recombinant Protein Polymer
TITLE OF INVENTION: Synthetic Protein Polymer
NUMBER OF SEQUENCES: 118
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: CA
COUNTRY: US
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/609,716
FILING DATE: 06-NOV-1990
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:

NAME: Rowland, Bettam I
REGISTRATION NUMBER: 20015
REFERENCE/DOCKET NUMBER: A-55186-3/BIR
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-781-1989
TELEFAX: 415-398-3249
INFORMATION FOR SEQ. ID NO: 66:
SEQUENCE CHARACTERISTICS:
LENGTH: 357 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-609-716-66

Query Match 25.3%; Score 560; DB 1; Length 357;
Best Local Similarity 38.8%; Pred. No. 7.7e-29;
Matches 131; Conservative 7; Mismatches 126; Indels 74; Gaps 12;

QY 48 AAPPLPPDPMPALNSALFAPNPVPPAPPGPNSAPAPMPPTPLPPVPPSGAPR-- 105
DB 24 AHPFFASDP-----MGAPGPPGPPGPPGPPGAPGPPGPPGPPGPPGPPGPP 74
QY 106 -----PVPAVP--PMPPAPKRMPLAPAPAPSPISLAVPVPPVPPVPLPVKM 155
DB 75 AGKGAHGPAPGPPGAPGPPGAPGPPGAPGPPGAPGPPGAPGPPGAPGPPGAP 125
QY 156 P-----PSPVPPPPAPDEPNPPAPAPPLPNSPAPPPVPPVPPV 198
DB 126 PGPKGDGADGAPKADGSPPPAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 185
QY 199 PLTLNPVPPAPAPANTSNSPLRPPAPAPPLKPPAPPMPPAPAPNSPAPSPSPVP 258
DB 186 PGCKDGRDAGPKGADGSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 243
QY 259 VFPTPPGPPAPPPPPNSPPAPAPAPAPAPPLPNSPAPAPPLPNSPAPPPAPGAP 318
DB 244 GLPGPKDGRDAGPKGADGSP--GPAGPVGPPAPGPPGPP-----GPPG--PGAPG 292
QY 319 PPAPPLYSPPAPAPCPVP-----GAPLAPLP 346
DB 293 PPGPP-----GPAGPPGLPKPKGDKGADGPKGADGSPGP 326

RESULT 9
US-08-642-255-33
Sequence 33, Application US/08642255
Patent No. 5773249

GENERAL INFORMATION:
APPLICANT: Cappello, Joseph
APPLICANT: Ferrari, Franco A.
TITLE OF INVENTION: High Molecular Weight Collagen-Like
TITLE OF INVENTION: Protein Polymers
NUMBER OF SEQUENCES: 135
CORRESPONDENCE ADDRESS:
ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/642,255
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: ROWLAND, Bettam I.

;; CURRENT APPLICATION NUMBER: US/09/219,849
;; CURRENT FILING DATE: 1998-12-23
;; NUMBER OF SEQ ID NOS: 50
;; SOFTWARE: Patent In Ver. 2.1
;; SEQ ID NO 5
;; LENGTH: 960
;; TYPE: PRT
;; ORGANISM: Artificial Sequence
;; FEATURE:
;; OTHER INFORMATION: Description of Artificial Sequence: Illustrative
;; OTHER INFORMATION: amino acid sequence
US-09-219-849-5

Query Match 24.4%; Score 540.5; DB 3; Length 960;
Best Local Similarity 34.9%; Pred. No. 3e-27;
Matches 164; Conservative 7; Mismatches 180; Indels 119; Gaps 24;

QY 7 PAPPAPSPINP-----PVPPVPLPAPARTLSPPVPPAPPSPISLAAPLPDPMPMPA 61
D 32 PGPPGAPGAPGPPGSRDPPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPG 87
QY 62 IMSALEARNPVPPA--PPGNSAPAR-----PMPTTLPPLPPVPPGSGAPPPVAPV 111
D 88 -----APGAPGPPGSRDPPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPG 141
QY 112 PMPPAPAPMPAPAPAPAPSPPTSWLAVPVPPVPPVPLPVKM-----PPSPVPPPPA 167
D 142 PAPPGPPGSRDPPG--PPGAPGAPGPPG--RDPPGAPGAPGAPGAPGAPGAPGAPGAPG 198
QY 168 --EETPPNPAPAPAPPLENSPPPPV-----PAPPPPPPPPPPPPPPPPPPPPPPP 191
D 199 SRDPPGPPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPG 258
QY 192 -PPVPPVPLTLNPVPPAPPAANT--SNSPLRPPA-----PAPPLKGPAP-- 237
D 259 GAPPAPAPGPPGSRDPPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPG 318
QY 238 ----PMPAPNSPAP-----PSPSPPPVPPF--TPGPPAPPEPNSPPA--PPAP 283
D 319 GSRDPPGPPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAP 377
QY 284 AAPLP--GSPAPAPAPPLPNSPAPPG--PPAMPGADPPAPPLPYSSPPAPACP---- 336
D 378 GAPPAPAPGPPGSRDPPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPG 435
QY 337 -----VGAFLAPLPISGRPSNSWGVFTMISPSNGAAAAAALAYAPA 381
D 436 PPGSRDPPGAPGAP--AGPPGSRDPPG-----PPGAPGAPGAPGAPGAP 477
b

RESULT 12
US-08-642-255-62
; Sequence 62, Application US/08642255
; Patent No. 5773249
; GENERAL INFORMATION:
; APPLICANT: CAPPELLO, Joseph
; APPLICANT: FERRARI, Franco A.
; TITLE OF INVENTION: High Molecular Weight Collagen-Like
; NUMBER OF SEQUENCES: 135
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30

;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/642,255
;; FILING DATE:
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: ROWLAND, Bertam I.
;; REGISTRATION NUMBER: 20,015
;; REFERENCE/DOCKET NUMBER: A55556-3/BIR
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (415) 494-8700
;; TELEFAX: (415) 494-8771
;; TELEEX: 910 277299 FHT UR
;; INFORMATION FOR SEQ ID NO: 62:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1064 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
US-08-642-255-62

Query Match 23.8%; Score 528.5; DB 1; Length 1064;
Best Local Similarity 32.9%; Pred. No. 1.8e-26;
Matches 150; Conservative 7; Mismatches 200; Indels 99; Gaps 17;

QY 7 PAPPAPSPINPVPVPPPLPAP--RTLSPPVPPAPPSPISLAAPLPDPMPMPAI 62
D 54 PGAPGPPGSRDPPGPPAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPG 113
QY 63 WSALEARNPVPPAPPGPNSAPAPMPPTPLPPVPPG--GAPPPVAVPMPAPKRM 121
D 114 PGSGDPPGPPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPG 173
QY 122 ALPPAPAPSPPTSWLAVPVPPVPPVPLPVKM-----PPSPVPPPP--AEETPN 173
D 174 PGPPGAPGAPGPPG--RGDPPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPG 232
QY 174 PPAAPAPPLENSPPPPVPPVPPVPLTLNPVPP-----PAPPPPPPPPPPPPPPPPP 209
D 233 APGAPGPPGSRDPPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAP 292
QY 210 PPAANTNSLRPAPAPPLK--PGPAPMPAPNSPAPSPVPPVPP-- 262
D 293 PKGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPG 352
QY 263 PGGPAPPEPNSPP--APPAPAPLPG-----PSPAPAPPLPNSPAPAPGPP 311
D 353 DPGPPGAPG--AGPPGSRDPPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPG 411
QY 312 AMPGA-----PDPPAPPLPPVSSP-----PAPPPAC-----VGAFLAPLP 347
D 412 GAPPAPGPPGSRDPPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPG 470
QY 348 SGRPSNSWGVFTMISPSNGAAAAAALAYAPA 383
D 471 AGPPG-----SRGDPPGPPG-----AGCPA 489

RESULT 13
US-08-642-255-132
; Sequence 132, Application US/08642255
; Patent No. 5773249
; GENERAL INFORMATION:
; APPLICANT: CAPPELLO, Joseph
; APPLICANT: FERRARI, Franco A.
; TITLE OF INVENTION: High Molecular Weight Collagen-Like
; NUMBER OF SEQUENCES: 135
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco

STATE: California
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/642,255
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Rowland, Bertram I.
REGISTRATION NUMBER: 20,015
REFERENCE/DOCKET NUMBER: A55556-3/BIR
TELEPHONE: (415) 494-8700
TELEFAX: (415) 494-8771
TELEX: 910 277299 FHT UR
INFORMATION FOR SEQ ID NO: 132:
SEQUENCE CHARACTERISTICS:
LENGTH: 829 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-642-255-132

Query Match 21.6%, Score 479.5; DB 1; Length 829;
Best Local Similarity 33.7%, Pred. No. 1.8e-23;
Matches 145; Conservative 19; Mismatches 191; Indels 75; Gaps 18;
QY 3 MPVPAPAPSPINPPVP-----VPLPAAPRTLS-----PVPPAPSPIS 46
DB 367 LPSGAGTGTGPGQLPSPGAGTGTGPGQLPSPGAGTGTGPGQLPSPGAGTGTGPG 426
QY 47 LAAPLPDPMPAPAIWSALEAPNPVPAPPGNSAPAPMPPTPLPVPPGSGARP 106
DB 427 LPSGAGTGTGPGQLPSPGAGTGTGPGQLPSPGAGTGTGPGQLPSPGAGTGTGPG 486
QY 107 VPAPVPM--PAPKMPALPPAP--PAPSPPTSLAVVPVPVPPVPLPVKMPSPVP 162
DB 487 LPSGAGTGTGPGQLPSPGAGTGTGPGQLPSPGAGTGTGPGQLPSPGAGTGTGPG 539
QY 163 --PPPPAPETPNPPAPAP--LENSPPPPVP-----PVPPVPLTLNPPVPPAPAN 214
DB 540 GTPGPGQLPSPGAGTGTGPGQLPSPGAGTGTGPGQLPSPGAGTGTGPGQLPSPGAG 599
QY 215 TSNSFLRPAPAPPLKGPAP--PPMPAPNSPAPSPSPVPVFPTR-----PGPP 267
DB 600 GTPGPGQLPSPGAGTGTGPGQLPSPGAGTGTGPGQLPSPGAGTGTGPGQLPSPGAP 659
QY 268 APPEPNSPPAPAP-----PAAP--LPGP-----SPAPAP--PLPNSPAP 307
DB 660 GTPGPGQLPSPGAGTGTGPGQLPSPGAGTGTGPGQLPSPGAGTGTGPGQLPSPGAP 719
QY 308 --PGPPAMPGADPPAPPLPYSP-----PAPPACP--VPGAPLAPLP--ISGRPSNS 354
DB 720 GTPGPGQLPSPGAGTGTGPGQLPSPGAGTGTGPGQLPSPGAGTGTGPGQLPSPGAM 779
QY 355 WGVGFTMLSR 364
DB 780 DPGRYHMAAK 789

RESULT 14
US-08-397-633A-53
Sequence 53, Application US/08397633A
Patent No. 5773577
GENERAL INFORMATION:
APPLICANT: Cappello, Joseph

TITLE OF INVENTION: PRO
TS COMPRISING SUBSTRATESCAPABLE
SYNTHATIC CROSS-LINKING
NUMBER OF SEQUENCES: 10;
CORRESPONDENCE ADDRESS:
ADDRESSEE: FLEHR, HOH
STREET: 4 Embarcadero
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy d
COMPUTER: IBM PC comp
OPERATING SYSTEM: PC-1
SOFTWARE: Patent Re
CURRENT APPLICATION DATA
APPLICATION NUMBER: U
FILING DATE:
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATI
NAME: Rowland, Bertra
REGISTRATION NUMBER:
REFERENCE/DOCKET NUMBE
TELEPHONE: (415) 781-
TELEFAX: (415) 398-32
TELEX: 910 277299
INFORMATION FOR SEQ ID NO:
SEQUENCE CHARACTERISTICS
LENGTH: 829 amino aci
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-397-633A-53

Query Match 21.6%, Score 479.5; DB 1; Length 829;
Best Local Similarity 33.7%, Pred. No. 1.8e-23;
Matches 145; Conservative 19; Mismatches 191; Indels 75; Gaps 18;
QY 3 MPVPAPAPSPINPPVP-----VPLPAAPRTLS-----PVPPAPSPIS 46
DB 367 LPSGAGTGTGPGQLPSPGAGTGTGPGQLPSPGAGTGTGPGQLPSPGAGTGTGPG 426
QY 47 LAAPLPDPMPAPAIWSALEAPNPVPAPPGNSAPAPMPPTPLPVPPGSGARP 106
DB 427 LPSGAGTGTGPGQLPSPGAGTGTGPGQLPSPGAGTGTGPGQLPSPGAGTGTGPG 486
QY 107 VPAPVPM--PAPKMPALPPAP--PAPSPPTSLAVVPVPVPPVPLPVKMPSPVP 162
DB 487 LPSGAGTGTGPGQLPSPGAGTGTGPGQLPSPGAGTGTGPGQLPSPGAGTGTGPG 539
QY 163 --PPPPAPETPNPPAPAP--LENSPPPPVP-----PVPPVPLTLNPPVPPAPAN 214
DB 540 GTPGPGQLPSPGAGTGTGPGQLPSPGAGTGTGPGQLPSPGAGTGTGPGQLPSPGAP 599
QY 215 TSNSFLRPAPAPPLKGPAP--PPMPAPNSPAPSPSPVPVFPTR-----PGPP 267
DB 600 GTPGPGQLPSPGAGTGTGPGQLPSPGAGTGTGPGQLPSPGAGTGTGPGQLPSPGAP 659
QY 268 APPEPNSPPAPAP-----PAAP--LPGP-----SPAPAP--PLPNSPAP 307
DB 660 GTPGPGQLPSPGAGTGTGPGQLPSPGAGTGTGPGQLPSPGAGTGTGPGQLPSPGAP 719
QY 308 --PGPPAMPGADPPAPPLPYSP-----PAPPACP--VPGAPLAPLP--ISGRPSNS 354
DB 720 GTPGPGQLPSPGAGTGTGPGQLPSPGAGTGTGPGQLPSPGAGTGTGPGQLPSPGAM 779
QY 355 WGVGFTMLSR 364
DB 780 DPGRYHMAAK 789

RESULT 15

US-08-175-155-68

; Sequence 68; Application US/08175155

; Patent No. 5641648

; GENERAL INFORMATION:

; APPLICANT: Ferrari, Franco A.

; APPLICANT: Cappello, Joseph

; APPLICANT: Crissman, John W.

; APPLICANT: Dorman, Mary A.

; TITLE OF INVENTION: Methods for Preparing Synthetic

; NUMBER OF SEQUENCES: 69

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Flehr, Hobach, Test, Albilton & Herbert

; STREET: Four Embarcadero Center, Suite 3400

; CITY: San Francisco

; STATE: CA

; COUNTRY: US

; ZIP: 94111

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/175,155

; FILING DATE: 29-DEC-1993

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Rowland, Bertram I.

; REGISTRATION NUMBER: 20015

; REFERENCE/DOCKET NUMBER: A-55186-5/BIR

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 415-781-1989

; TELEFAX: 415-398-3249

; INFORMATION FOR SEQ ID NO: 68:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 837 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

; US-08-175-155-68

Query Match

21.6%; Score 478; DB 1; Length 837;

Best Local Similarity 34.6%; Pred. No. 2.2e-23;

Matches 140; Conservative 17; Mismatches 178; Indels 70; Gaps 17;

QY 3 MPVPAPAPAPSPINPVPP-----VPLPAPRTLSP-----PVPPAPPSPI 46

Db 44 LPSGPPGAPGTPGGGLGSPGAPGTPGGGLGSPGAPGTPGGGLGSPGAPGTPGG 103

QY 47 LAAPLPPDPMPPAIWSALEAPNPVPAPAPGPNASAPAPMPPTPLPVPVPGSGAPR 106

Db 104 LPSGPPGAPGTPGGGLGSPGAPGTPGGGLGSPGAPGTPGGGLGSPGAPGTPGG 163

QY 107 VPAVPMP--PAPKMPALPPAP--PAPSPPTSMILAVVPVPPVPPPLPVKMPSPVP 162

Db 164 LPSGPPGAPGTPGGGLGSPGAPGTPGGGLGSP-----PGAPGTPGGGLGSPGAP 216

QY 163 --PPPPAEPETPNPAPAPAP--LENSPPPPVP-----PVVPVPLTLNPPVPPAPPA 214

Db 217 GTPGPPGGLPSPGAPGTPGGGLGSPGAPGTPGGGLPSPGAPGTPGGGLGSPGAP 276

QY 215 TSNPLRPAPAPAPLKPGRPA--PPMPAPNSPAPAPSPSPVPVPTP-----RGPP 267

Db 277 GTPGPPGGLPSPGAPGTPGGGLGSPGAPGTPGGGLPSPGAPGTPGGGLGSPGAP 336

QY 268 APPEPNSPPAPAP-----PAAP-LPGP-----SPAPAPAP-----PLPNSPAP 307

Db 337 GTPGPPGGLPSPGAPGTPGGGLGSPGAPGTPGGGLPSPGAPGTPGGGLGSPGAP 396

QY 308 --PGPPAPGAPDPAPAPLIPYSSP--PAPPAC-----VPGAPLAP 344

Db 397 GTPGPPGGLPSPGAPGTPGGGLGSPGAPGTPGGGLPSPGAP 441

Search completed: January 5, 2001, 14:27:39

Job time: 231 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 5, 2001, 14:26:45 ; Search time 62.98 Seconds
(without alignments)
209.571 Million cell updates/sec

Title: US-09-461-774-10
Perfect score: 2216
Sequence: 1 NSMPVPAPAPAPPPINPPV.....NGAAMAAASALVAPAPAVKV 386

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 268485 seqs, 34193795 residues
Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: A.Geneseq_36.*
2: /SIDSI/gcgdata/geneseq/geneseqp/AA1980.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseqp/AA1981.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseqp/AA1982.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseqp/AA1983.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseqp/AA1984.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseqp/AA1985.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseqp/AA1986.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseqp/AA1987.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseqp/AA1988.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseqp/AA1989.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseqp/AA1990.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseqp/AA1991.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseqp/AA1992.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseqp/AA1993.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseqp/AA1994.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseqp/AA1995.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseqp/AA1996.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseqp/AA1997.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseqp/AA1998.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseqp/AA1999.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseqp/AA2000.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1065.5	48.1	572	18 W31855	Mycobacterium tube
2	1065.5	48.1	763	18 W31852	Mycobacterium tube
3	832	37.5	408	17 W07539	Collagen-like prot
4	721	32.5	330	19 W57645	Collagen-like poly
5	598.5	27.0	561	14 R37739	Collagen-like poly
6	598.5	27.0	561	17 R93249	Collagen-like poly
7	598.5	27.0	561	19 W57650	Collagen-like poly
8	570.5	25.7	777	14 R37740	Collagen-like poly
9	570.5	25.7	777	17 R93250	Collagen-like poly
10	570.5	25.7	777	19 W57651	Collagen-like poly
11	560	25.3	357	17 R95115	Intervening sequen
12	560	25.3	357	19 W57646	Collagen-like poly

13	528.5	23.8	1064	17 R93254	Collagen-like poly
14	528.5	23.8	1064	19 W57652	Collagen-like poly
15	528.5	23.8	1065	14 R37741	Collagen-like poly
16	519.5	23.4	783	19 W37151	Mouse neural Mena+
17	519.5	23.4	787	19 W37152	Mouse neural Mena+
18	519.5	23.4	802	19 W37153	Mouse neural Mena+
19	490	22.1	1404	13 R26049	MSF precursor. Sy
20	479.5	21.6	829	16 R80334	Protein polymetric
21	479.5	21.6	829	19 W49723	Protein polymer ad
22	479.5	21.6	829	19 W57673	Collagen-like poly
23	478	21.6	837	16 R80320	CLP 3.7 protein mu
24	478	21.6	837	18 W26353	Collagen-like prot
25	478	21.6	837	19 W57659	Collagen-like poly
26	478	21.6	837	21 W53530	Amino acid sequenc
27	478	21.6	837	21 W78290	CLP 3.7 amino acid
28	472	21.3	417	18 W26354	Collagen-like poly
29	472	21.3	417	19 W57660	Collagen-like poly
30	472	21.3	417	19 W53531	Amino acid sequenc
31	472	21.3	417	21 W78291	CLP 3.7 amino acid
32	457.5	20.6	252	14 R37738	Collagen-like poly
33	457.5	20.6	252	17 R93248	Collagen-like poly
34	454	20.5	504	20 R24091	Human Wiskott-Aldr
35	452.5	20.4	378	12 R14162	PRP encoded by clo
36	445	20.1	3119	19 W72204	HSV-2 strain S85 C
37	441.5	19.9	378	12 R14160	PRP 378. Trilicium
38	432	19.5	1664	19 W43106	C. thermocellum O1
39	431.5	19.5	682	16 R80330	Protein polymetric
40	431.5	19.5	682	19 W49717	Collagen-like poly
41	431.5	19.5	682	19 W57670	Collagen-like poly
42	427.5	19.3	174	19 W81727	M. tuberculosis im
43	427.5	19.3	174	20 W64360	Mycobacterium tube
44	427.5	19.3	174	20 Y39157	M. tuberculosis an
45	427.5	19.3	174	20 Y39014	M. tuberculosis re

ALIGNMENTS

RESULT	ID	ALIGNMENTS
1	W31855	W31855 standard; Protein; 572 AA.
XX	W31855;	
AC	W31855;	
XX		
DT	27-APR-1998 (first entry)	
DE	Mycobacterium tuberculosis 55 kDa protein.	
XX		
KM	Tuberculosis; mycobacteria; infection; diagnosis;	
KW	antimycobacterial; antibiotic; vaccine.	
OS	Mycobacterium tuberculosis.	
PN	W09741252-A2.	
XX		
PD	06-NOV-1997.	
XX		
PF	18-APR-1997; 97WO-EP01973.	
XX		
PR	29-APR-1996; 96DE-4017184.	
PA	(GBF) GBF GES BIOTECH FORSCHUNG GMBH.	
XX		
PI	Esplitia C, Honisch C, Moreno C, Singh M;	
XX		
DR	WPI; 1997-549750/50.	
XX		
DR	N-PDSB; T93610.	
XX		
PT	New DNA and related proteins or RNA derived from M. tuberculosis -	
XX	used for diagnosis of mycobacterial infections, monitoring	
XX	vaccination and development of anti-mycobacterial agents	
PS	Claim 11; Fig 16; 55pp; English.	

XX This novel 55 kDa protein is encoded by an open reading frame of
 CC a Mycobacterium tuberculosis DNA fragment (see T93610) containing
 CC polymorphic GC-rich sequences. Its amino acid sequence shows
 CC a high proline content, but there is no homology to any known
 CC proline-rich antigens of Mycobacterium. Novel M. tuberculosis
 CC proteins (see W31851-57) are claimed. These can be produced as
 CC recombinant proteins, especially in bacterial, yeast, fungal or
 CC higher eukaryote host cells, and used for diagnosing tuberculosis
 CC and other mycobacterial infections in humans or animals. The
 CC claimed proteins can also be used for epidemiological studies, for
 CC monitoring vaccination, and for the development of vaccines and
 CC anti-mycobacterial drugs.

Sequence 572 AA:

Query Match 48.1%; Score 1065.5; DB 18; Length 572;
 Best Local Similarity 42.3%; Fred. No. 5e-52; Mismatches 91; Indels 193; Gaps 24;
 Matches 233; Conservative 31;

4 PVPAPAPAPSPFINDPVPVPPPLPAAPRTLSPPVPPAPSPISLAA-----PPIPP--- 54
 44 PAPPAPPIPPAPV--PIPPVPIPIPPVPIKI--PPAPPAP--PVAAVAVIAPCPPIPIPI 98
 55 -DPPMPA-----IWS----- 64
 99 nhppapapavpyvplaplpnshppapsapvyvplaplpisgrprvawkgstlctf 158
 65 -----ALEAPNPVPPAPPGP-NSA--PAPMP-- 90
 159 ccrvcsgevlagalnpsrpsrslttlcpalpaplpplppllnlavppilpplvt 218
 91 -----TPPLPVPVPGSGARPRVAVPPMPAPKRMPLPPA----- 126
 219 alappilpplaplpispyvppaplpipgk--pwltcpplapapppctvplppgpcpse 276
 127 ---PPAPSPPTSMILAVVPVPPVPPPLP--VKMPSPVPVPPAPPEPT-----P 172
 277 knpppapppepkspalppappppsavrvppspilppapaprasmpalppap 336
 173 NPRA-----PPAPPLNSPPPPVPPV-----PPVPLTLNPVPPAPPA-- 212
 337 sppatrlcpplppspapppapppapptpkllsanppccvppapnrvppapppap 396
 213 -----ANTSNSPLRPAPAPPLK---PGPPAPMPAPNS-----PAAPSPSP 255
 397 elpappdpptcpvansppapppapppsalpfvnpappptcpaaksrpalpaappap 456
 256 PVPVFPPTPPGAPAPPEPNS--PPAPAPPAAPLPGSPAPADPDLNSPAPPGPPAP 313
 457 pvrattcpappapppasmlalppapdpplppl--atcpappapplmsppapplp 513
 314 PGADPPAPPLPYSSPPAPACVPGAPLAPLPISGR-----SNSWGVFTLSPRSNGA 369
 514 --apppappltpngppspilapvpgablplpigrvfarfkna1g-----sssgd 564
 370 AAASALA 377
 565 taasaaa 572

RESULT 2
 W31852 standard; protein; 763 AA.
 W31852:
 27-APR-1998 (first entry)
 Mycobacterium tuberculosis 74 kDa protein.
 Tuberculosis; mycobacteria; infection; diagnosis;

KW antimycobacterial; antib
 OS Mycobacterium tuberculo
 PN WO9741252-A2.
 PD 06-NOV-1997.
 XX 18-APR-1997; 97WO-EP01
 XX 29-APR-1996; 96DE-4017
 PA (GBFB) GBF GES BIOTEC
 XX Esplita C, Honisch C,
 DR WPI: 1997-549750/50.
 DR N-PSDB: T93610.
 XX New DNA and related prot
 PT used for diagnosis of my
 PT vaccination and developm
 XX Claim 5; Fig 13; 55pp; E
 XX This novel 74 kDa protei
 CC a Mycobacterium tuberculo
 CC polymorphic GC-rich sequ
 CC a high proline content,
 CC proline-rich antigens of
 CC proteins (see W31851-57)
 CC recombinant proteins, es
 CC higher eukaryote host ce
 CC and other mycobacterial
 CC claimed proteins can als
 CC monitoring vaccination,
 CC anti-mycobacterial drugs

Query Match 48.1%; Score 1065.5; DB 18; Length 763;
 Best Local Similarity 42.3%; Fred. No. 6.2e-52; Mismatches 91; Indels 193; Gaps 24;
 Matches 233; Conservative 31;

4 PVPAPAPAPSPFINDPVPVPPPLPAAPRTLSPPVPPAPSPISLAA-----PPIPP--- 54
 235 PAPPAPPIPPAPV--PIPPVPIPIPPVPIKI--PPAPPAP--PVAAVAVIAPCPPIPIPI 289
 55 -DPPMPA-----IWS----- 64
 290 nhppapapavpyvplaplpnshppapsapvyvplaplpisgrprvawkgstlctf 349
 65 -----ALEAPNPVPPAPPGP-NSA--PAPMP-- 90
 350 ccrvcsgevlagalnpsrpsrslttlcpalpaplpplppllnlavppilpplvt 409
 91 -----TPPLPVPVPGSGARPRVAVPPMPAPKRMPLPPA-----PPIPP--- 126
 410 alappilpplaplpispyvppappppsavrvppspilppapaprasmpalppap 467
 127 ---PPAPSPPTSMILAVVPVPPVPPPLP--VKMPSPVPVPPAPPEPT-----P 172
 468 knpppapppepkspalppappppsavrvppspilppapaprasmpalppap 527
 173 NPRA-----PPAPPLNSPPPPVPPV-----PPVPLTLNPVPPAPPA-- 212
 528 sppatrlcpplppspapppapppapptpkllsanppccvppapnrvppapppap 587
 213 -----ANTSNSPLRPAPAPPLK---PGPPAPMPAPNS-----PAAPSPSP 255
 588 elpappdpptcpvansppapppsalpfvnpappptcpaaksrpalpaappap 647

	Query Match	37.5%;	Score 832;	DB 17;	Length 408;	
	Best Local Similarity	46.8%;	Pred. No. 2.5e-39;			
	Matches 169;	Conservative 5;	Mismatches 155;	Indels 32;	Gaps 11;	
OY	4 PVPPAPAPSPSINPVVPVPLPA-----APRTLSPVPVAPPSPISLAAPPILPDDPMF	59				
Dd	42 ppppppppqagppppppppppppgagvsgagagppppppppppppppppgagpppppppppp	101				
OY	60 PAIMSAI---EADNPVPVAPPGGNSAAPAPMPPTTPLLVPVPGGS-GAF-RVPAVPPMP	114				
Dd	102 pqpaqpvgsqagappppppppppgagadqpqqppppppgagavsgagadqpqpqpqr	161				
OY	115 PAKRMKMLPAPAPAPPSPTSMILAVPVVPVPPVPLPYKKMPESPVPVPPFFPAPEETPMR	174				
Dd	162 ppppgagap-pppppppppppp-----gpagvsgagagppppppppppppppgagpppp	212				
OY	175 PAPAPAPLSENSP-----PPPPVPVPVPVPLTLNLPVPVAPPAPANTSNSRLRPAPAPAP	228				
Dd	213 pppppppppagagvsgagagppppppppppppppgagppppppppppppppgagvsgagppqr	272				
OY	229 PLKCPAPAPMPMPAPANSAPAPSPSPVPVPTTPCGAPAPPEPNSSPPAPAPPAALPR	288				
Dd	273 pppppppppppgagppppppppppgagqv--gspgagppp-----gpqpqpqpqpagppr	326				
OY	289 GPSAPAPAPAPPLPNSPAAAPCGPAMPGAPDPPAPPLPVSPPAPAPACPVGAPLAPLPIIS	348				
Dd	327 -ppppppppppppgagvsgagppppppppppppppp-gagppppppppppppgagvsgp	384				
OY	349 G 349					
Dd	385 g 385					
RESULT	4					
ID	W57645					
XX	W57645 standard; peptide; 330 AA.					
AC	W57645:					
DT	27-AUG-1998 (first entry)					
DE	Collagen-Like polymer.					
KW	Collagen-like polymer; synthetic polymer; fibre coating;					
OS	prothetic device; catalytic substance.					
XX	Synthetic.					
PN	US5773249-A.					
XX	W57645					
ED	30-JUN-1998.					
XX	02-MAY-1996;	96US-0642255.				
XX	02-MAY-1996;	96US-0642255.				
PR	04-NOV-1986;	86US-0927258.				
PR	29-OCT-1987;	87US-0114618.				
PR	09-NOV-1988;	88US-0269429.				
PR	06-NOV-1990;	90US-0609716.				
PR	12-NOV-1991;	91US-0791960.				
PR	05-NOV-1992;	92US-0972032.				
PR	22-DEC-1995;	95US-0577046.				
PA	(PROT-) PROTEIN POLYMER TECHNOLOGIES INC.					
PI	Cappelletto J, Ferrari FA;					
DR	WPI, 1998-387004/33.					
XX	Recombinant collagen-like polymers - useful for making gels, films,					
PT	fibres, etc.					
XX						

ID	R93249 standard; Protein; 561 AA.
XX	
AC	R93249;
XX	
DT	24-FEB-1997 (first entry)
XX	
DE	Collagen-like polymer sequence D gene 1 polymer protein (pP0166).
XX	
KW	collagen; repetitive triad motif; recombinant production;
KW	photographic; medical; structural; fibre.
XX	
OS	Synthetic.
XX	
PN	US5496712-A.
XX	
PD	05-MAR-1996.
XX	
PF	06-NOV-1990; 90US-0609716.
XX	
PR	05-NOV-1992; 92US-0972032.
PR	06-NOV-1990; 90US-0609716.
PR	12-NOV-1991; 91US-0791960.
PA	(PROT-) PROTEIN POLYMER.
XX	
PI	Cappello J, Ferrari FA;
XX	
DR	WPI; 1996-150728/15.
XX	
PT	Collagen-like polymers comprising repetitive triads - produced in
XX	unicellular organisms with improved characteristics, useful in, e.g.
PT	photographic and medical fibres
XX	
PS	Example 2; Column 18; 43pp; English.
XX	
CC	The invention concerns collagen-like polymers having repetitive triads
CC	with reduced proline content, and where glycine is the initial amino
CC	acid and the subsequent amino acids are varied. The choice of triads
CC	utilised in a recombinant collagen-like polymer are chosen in order to
CC	affect properties such as helix stability, hydration, solubility, gel
CC	point, biodegradation and immunogenicity. Triads of particular interest
CC	include GAP, GPA, GPP, GAS, GPG, GPS, GPO, GSP, GLO, GPR, GPK, GAR,
CC	GER, GDR, GEP, GDA, GAH and GEA. The collagen-like polymers may impart
CC	new characteristics, finding wide use in photographic, medical,
CC	structural and fibre applications, and are capable of being produced in
CC	unicellular microorganisms at high mol. wts. and in high efficiency.
CC	The present sequence, encoded by pP0166, is a sequenced gene 1 polymer
CC	prepn. of antibodies.
XX	
Sequence	561 AA;

	Query Match	27.0%	Score 598.5	DB 17	Length 561	
	Best Local Similarity	35.6%	Pred No. 2e-26			
	Matches 154	Conservative	3	Mismatches 200	Indels 75	Gaps 14
Oy	4 PVPAPAPADPSINPVPVPP-LPAAPTLSPPVPAPDPSISLAAPLPDPEPMAPAI	62				
Dd	54 pppapppagppagppgagppgagppgagppgagppgagppgagppgagppagp	113				
Oy	63 WSLALPNPVPVPPAPPGMSARAPMHPTRPLLPVPRGSGAP-----RVPAVPPMAP	117				
Dd	114 pgapgpaggpgagppgagppgagppgagppgagppgagppgagppgagppgagpp	173				
Oy	118 KRM-----PALPAPAPAPSPPTSWLAVNPVPVPPPLP-----VKMPSPDPVFPPA	167				
Dd	174 agppgagppagppgagppgagpp-----gagppgagppgagppgagppgagppgagppag	229				
Oy	168 EETENPPAPAPAPLENSSPPPVPPVPLTLNP-----V	206				
Dd	230 ppgapppagppagppagppgagppgagppgagppgagppahpapgkghagppgkghag	289				

OY	207	PPAPPAANTSNPLRPAPAPAPLKEGPPAPMPKMAPNSPAAPESPSP-----VPV	259
Dd	230	paggkgaigpaqgkagapppagppagppagppagppagppagppagppagppagppag	349
OY	260	FPTPPGPPAPPEPNSSP--PAPPAPPAALP-----GPSPPAPPAAPPLPNSPAAPECPA	312
Dd	350	apgpaqpppgagppagpppgagppagppagppagppagppagppagppagppagppag	409
OY	313	WGPADPPAPLPYSPPAPACVP- GAFIADLPISGRPSNWGVFTMLSRPSNGAA	371
Dd	410	ppgagppagpp- gappgagppgagppagppagpp- agpp-----gagp	451
OY	372	AASALATAPAA	383
Dd	452	pagppg-agpqa	462
RESULT	7		
ID	M57650	M57650 standard; peptide; 561 AA.	
XX	AC	M57650;	
XX	DE	27-AUG-1998 (first entry)	
XX	XX	Collagen-like polymer.	
KW	XX	Collagen-like polymer; synthetic polymer; fibre coating;	
XX	XX	prosthetic device; catalytic substance.	
OS	XX	Synthetic.	
XX	PN	US5773249-A.	
PD	30-JUN-1998.		
XX	PF	02-MAY-1996; 96US-0642255.	
XX	PR	02-MAY-1996; 96US-0642255.	
XX	PR	04-NOV-1986; 86US-0927258.	
PR	29-OCT-1987; 87US-0114618.		
PR	09-NOV-1988; 88US-0269429.		
PR	06-NOV-1990; 90US-0609716.		
PR	12-NOV-1991; 91US-0791960.		
PR	05-NOV-1992; 92US-0972032.		
PR	22-DEC-1995; 95US-0577046.		
PA	(PROT-) PROTEIN POLYMER TECHNOLOGIES INC.		
P1	Cappelletto J, Ferrari FA;		
DR	WPI; 1998-387004/33.		
XX	Recombinant collagen-like polymers - useful for making gels, films,		
PT	fibres, etc.		
XX	Example 3; Column 28; 93pp; English.		
PS	This sequence represents a unnatural collagen-like polymer		
CC	of the invention. The products may be used as films, fibres, moulded		
CC	objects and admixed with other natural or synthetic polymers or coatings		
CC	on fibres, films, labware or other surfaces, e.g. prosthetic devices. The		
CC	polymers may be used for binding a wide variety of specific binding		
CC	materials, as catalytic substances (where the amino acid sequence may		
CC	specifically chelate a wide variety of elements), as purification media,		
CC	composites, laminates or adhesives. They may also be combined with		
CC	inorganic or organic materials such as carbon fibres, nylon fibres,		
CC	nitracellulose, etc., as flask coatings or in synthetic matrices for the		
CC	growth and study of cells, as affinity columns or as supports for		
CC	biological materials. The polymers have collagen-like properties, but may		
CC	be easily expressed in micro-organisms in high efficiency. The new		
XX	sequences can be tailored to give the desired properties.		

[illegible]

```

OY 263 PPAPAPPEPSSSP--PAPAPPAALP-----GSPAPAPAPALPLNSPAAAPGPPAMPG 315
    ||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 333 PAPPAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAP 412
OY 316 APPDPAPLPYSSSPAPAPACVP-GAPLAPLPISGRPSMSWGVFTMLSRPSNGAANAAS 374
    ||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 413 aPPGAPGP-gAPPgAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPG 464
    ||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 375 ALAYAPA 381
    | : ||
Db 465 KgaNgpa 471

RESULT 10
W57651
W57651 standard; peptide; 777 AA.
XX
AC W57651;
XX
DT 27-AUG-1998 (first entry)
XX
DE Collagen-like polymer.
XX
KW Collagen-like polymer; synthetic polymer; fibre coating;
XX prosthetic device; catalytic substance.
OS Synthetic.
XX
PN US5773249-A.
PD 30-JUN-1998.
XX
PF 02-MAY-1996; 96US-0642255.
XX
PR 02-MAY-1996; 96US-0642255.
PR 04-NOV-1986; 96US-0642255.
PR 29-OCT-1987; 87US-0114618.
PR 09-NOV-1988; 88US-0269429.
PR 06-NOV-1990; 90US-0609716.
PR 12-NOV-1991; 91US-0791960.
PR 05-NOV-1992; 92US-0972032.
PR 22-DEC-1995; 95US-0577046.
XX
PA (PROT-) PROTEIN POLYMER TECHNOLOGIES INC.
PI Cappelletto J, Ferrari FA;
DR WPI; 1998-387004/33.
XX
PT Recombinant collagen-like polymers - useful for making gels, films,
XX fibres, etc.
XX
PS Example 3; Column 29; 93pp; English.
XX
XX This sequence represents a unnatural collagen-like polymer
of the invention. The products may be used as films, fibres, moulded
objects and admixed with other natural or synthetic polymers or coatings
on fibres, films, labware or other surfaces, e.g. prosthetic devices. The
CC polymers may be used for binding a wide variety of specific binding
materials, as catalytic substances (where the amino acid sequence may
CC specifically chelate a wide variety of elements), as purification media,
CC composites, laminates or adhesives. They may also be combined with
inorganic or organic materials such as carbon fibres, nylon fibres,
CC nitrocellulose, etc., as flask coatings or in synthetic matrices for the
growth and study of cells, as affinity columns or as supports for
CC biological materials. The polymers have collagen-like properties, but may
be easily expressed in micro-organisms in high efficiency. The new
sequences can be tailored to give the desired properties.
XX
XX Sequence 777 AA;
SO

```

Query Match 25.7% Score 570.5; DB 19; Length 777;
Best Local Similarity 34.0%; Pred. No. 8,5e-25;
Matches 145; Conservative 6; Mismatches 209; Indels 67; Gaps 11;

```
OY 4 PVPPAPAPPSINPVPPVPP-LPAAPRTLSPPVPPAPPSITSLAAPLPDPMPAPAI 62
DB 63 PGPAPGPAPGPAPGPAPGPAPGPAPGPAPGPAPGPAPGPAPGPAPGPAPGPAPGP 122
OY 63 WSALEAPVPPAPAPGPAPGPAPGPAPGPAPGPAPGPAPGPAPGPAPGPAPGPAPGP 112
DB 123 PGPAPGPAPGPAPGPAPGPAPGPAPGPAPGPAPGPAPGPAPGPAPGPAPGPAPGP 182
OY 113 MPAPAKRMPALPPAPAPAPSPPTSMILAVPPVPPVPPPLPVKMPSPVPPPPAPETP 172
DB 183 agpkghapagpkghapagppp-----gapgagppgagp-----gpagppgagppgagp 234
OY 173 NPAPAPAPLENSPPPPVPPVPPVPPPLTNPPVPPAPAPANTSNPLRPPAPAPLKP 232
DB 235 gpagppgagppgagppgagppgagppgagppgagppgagppgagppgagppgagpp 294
OY 233 GPAPAPMPAPNSPAP-----PSPSPVPPVPPPT 262
DB 235 gpagppgagppgagppgagpkghapagpkghapagpkghapagppgagppgagp--pg 352
OY 263 PPGPAPAPPEPNSP--PAPAPAPALP-----GSPAPAPAPPLPNSPAPPGAPAPMG 315
DB 353 PGPAPGPAPGPAPGPAPGPAPGPAPGPAPGPAPGPAPGPAPGPAPGPAPGPAPGP 412
OY 316 AEDPAPAPLVSSPPAPAPACVP--GAPLAPLPSGRPSNSWVETMLSRPSNGAAAAAS 374
DB 413 agpgagpp--gagpgagppgagppgagppgagppgagppgagppgagp-----pkghapagp 464
OY 375 ALAYAPA 381
DB 465 kgahgpa 471
```

RESULT 11
R95115
ID R95115 standard; Protein; 357 AA.

AC R95115;
DT 03-FEB-1997 (first entry)
DE Intervening sequence conty. collagen like protein (CLP)-CB.
KW Polymer; repeat unit; natural collagen; intervening oligopeptide;
file; film; membrane; emulsion; coating; collagen like protein;
specific binding material; catalyst; purification agent; composite;
laminar; adhesive; cell growth surface; affinity column;
biological material support; wound dressing; in vivo prothesis.

OS Synthetic.
XX US5514581-A.
XX PD 07-MAY-1996.
XX PF 04-NOV-1986; 86US-0927258.
XX PR 06-NOV-1990; 90US-0609716;
PR 04-NOV-1986; 86US-0927258;
PR 29-OCT-1987; 87US-0114618;
PR 09-NOV-1988; 88US-0269429;
PR 07-NOV-1989; 89WO-US05016.

XX (PROT-) PROTEIN POLYMER TECHNOLOGIES INC.

XX Cappelletto J, Ferrari FA;

XX WPI; 1996-238772/24.

PT DNA encoding protein com
PT linked by oligopeptide w
PT wound dressings

PS Example 3; Columns 101-1
XX
CC Novel DNA sequence encod
CC repeating units of 3-9 a
CC present intervening sequ
CC able to assemble into a
CC The polymer comprises at
CC intervening oligopeptide
CC The polymer can be used
CC emulsions, coatings, etc
CC adhesives, catalysts, pu
CC biological materials. Ty
CC dressings, and in vivo p
CC with good mechanical pro
CC can provide a ligand for
CC chemically reactive site

repeated fibroin derived segments -
cell adhesion properties useful, e.g. in
71pp; English.
a polymer comprising segments of
acids from natural collagen, i.e. the
conty. collagen like protein (CLP)-CB,
ed structures formable into articles
ast 2 segments joined by an unaligned
ther than the repeating unit.
make fibres, films, membranes,
useful as, e.g. specific binding
fices, agents, composites, laminates,
al applications include wound
theses. The polymer produces articles
ties, and the intervening oligopeptide
inding a mol., antibody, etc., or a
or coupling to proteins, etc..

Query Match 25.
Best Local Similarity 38.
Matches 131; Conservative

Score 560; DB 17; Length 357;
Pred. No. 1.8e-24;
7; Mismatches 126; Indels 74; Gaps 12;

```
OY 48 AAPPLPDPMPAPATWSAL 105
DB 24 ahpfasdp-----m 74
OY 106 -----PVAPV-PMP 155
DB 75 agpkgadgspgagpvgs 125
OY 156 P-----P 198
DB 126 pgpkdrgdaggkagdsp 185
OY 199 PLTNPPVPPAPAPANTSN 258
DB 186 pgpkdrgdaggkagdsp 243
OY 259 VFPTPPGPAPPEPNSPP 318
DB 244 g1pgkagdrdagpkagag 292
OY 319 PPAPPLPVSSPPAPACPV 346
DB 293 PPgpp-----gppgppg1pg 326
```

RESULT 12
W57646
ID W57646 standard; peptide 357 AA.

AC W57646;
XX 27-AUG-1998 (first ent
XX Collagen-like polymer.
XX Collagen-like polymer;
XX Collagen-like polymer;
XX prosthetic device; catal
XX Synthetic.

XX US5773249-A.

XX 30-JUN-1998.

XX Cappelletto J, Ferrari FA;
XX WPI; 1993-182496/22.
DR
XX
PT High mol. wt. collagen-like protein polymers - capable of being
PT produced in unicellular microorganisms
XX
PS Disclosure; Page 48; 82pp; English.
XX
CC This sequence is an example of a recombinantly produced DCP3 collagen-
CC like polymer (CLP) which consist of repeated tripeptide sequences
CC selected from a wide range of GXY sequences, where X and Y can be any
CC amino acid. The DNA encoding this sequence was cloned into plasmid
CC pPT 0176 and used to transform E. coli. DCP peptides comprise repeated
CC units of: A = GAGPAGP, B = GSRGPGP and/or C = GAGPAGP. These
CC polymers have molecular weights of >30 kD and are able to form helices
CC due to interchain linkages. These polymers pref. contain a proportion
CC of tripeptide triad sequences found in natural collagens, pref.
CC mammalian collagens. The CLPs impart unique characteristics to
CC materials such as fibres, membranes, films, coatings, hydrogels,
CC colloid suspensions and moulded articles.
XX
XX Sequence 1065 AA;
SQ

Query Match 23.8%; Score 528.5; DB 14; Length 1065;
Best Local Similarity 32.9%; Pred. No. 2.1e-22;
Matches 150; Conservative 7; Mismatches 200; Indels 99; Gaps 17;
QY 7 PAPPAPSPINPVVPPVPLPAP-----RTLSPPVPPAPSPISLAAPLPDDPMPPAI 62
Db 54 ppgpaggppsgdppppagppagppsgdppppagppagppsgdppppagppagp 113
QY 63 WSALEAPNPVPPAPPGPNSAPAPMPPTPLPVPVPGS-GAPRPVAPVPPMPAPKRM 121
Db 114 pgsrgdppppagppagppsgdppppagppagppsgdppppagppagppsgd 173
QY 122 ALPPAPAPSPPTSMALAVVPPVPPVPLPVKM-----PSPRPVPPFP---AEPETPN 173
Db 174 ppppaggppagppgs-rqdpaggppagppagppsgdppppagppagppsgdpp 232
QY 174 PPAAPPAPLENSPPPPVPPVPPVPLTLNPVPP-----A 209
Db 233 apgpaggppsgdppppagppagppsgdppppagppagppsgdppppagppag 292
QY 210 PPAANTSNLPPAPAPPLK---PGPAPVMPAPNSPAPSPSPVVPVFT--- 262
Db 293 pkghagppagppagppsgdppppagppagppsgdppppagppagppsg 352
QY 263 PPGPPAPPEPNSPP---APPAPPAALPG-----PSPAPAPPLPNSPAPAPGPP 311
Db 353 dppppagppp-agppgsrgdppppagppagppsgdppppagppagppsgdpp 411
QY 312 AMFGA-----PDPAPPLPYSSP-----PAPPACP-----VPGLAPLPI 347
Db 412 gagppagppsgdppppagppagppsgdppppagppagppsgdppppagppag 470
QY 348 SGRPSNSWGVFTMLSRPSNGAAAAALAYAPAPA 383
Db 471 agppg-----strgdppppg-----agppa 489

THIS PAGE BLANK (USPTO)